JESK J Ergon Soc Korea 2024; 43(4): 297-318
https://dx.doi.org/10.5143/JESK.2024.43.4.297
https://jesk.or.kr eISSN:2093-8462

Research on User Interface of Conversational AI for Foreign Language Speaking

Ga Young Yoo[1†], Seo Yeon Kim[2†], Kyungdoh Kim[3]

[1]Hongik University, Department of ART Studies, Seoul, 04066
[2]Hongik University, Department of Visual Communication Design, Seoul, 04066
[3]Hongik University, Department of Industrial & Data Engineering, Seoul, 04066

외국어 회화 학습을 위한 대화형 AI의 사용자 인터페이스 연구

유가영[1†], 김서연[2†], 김경도[3]

[1]홍익대학교 예술학과
[2]홍익대학교 시각디자인과
[3]홍익대학교 산업데이터공학과

Corresponding Author
Kyungdoh Kim
Hongik University, Department of Industrial & Data Engineering, Seoul, 04066
Email : kyungdoh.kim@hongik.ac.kr

Received : May 13, 2024
Revised : July 05, 2024
Accepted : August 12, 2024

† G. Yoo and S. Kim should be considered the co-first authors.

Copyright@2024 by Ergonomics Society of Korea. All right reserved.

Objective: This study proposes a user interface for designing a conversational AI for foreign language conversation learning, aiming to induce immersion and foster sustained user engagement.

Background: Conversational AI is gaining attention in the field of education for its potential to provide personalized learning and real-time feedback. In foreign language conversation learning through conversational AI, interaction with tutors plays a crucial role in influencing learners' immersion and intention to continue. Thus, consideration of anthropomorphic elements that can enhance such interaction is necessary. Anthropomorphism encompasses linguistic elements resembling human language (natural language models, human-like voices, etc.), non-verbal elements (emojis, meme usage, etc.), and visual elements (character images, profiles, etc.), all of which contribute to conveying a sense of social presence. While research integrating anthropomorphic elements into conversational AI is prevalent, there is still a lack of studies specifically focusing on anthropomorphic elements in the context of foreign language conversation learning. Therefore, this study proposes a user interface for conversational AI considering not only human-like appearance but also interaction-related elements such as linguistic and non-verbal factors.

Method: The study focuses on analyzing how the anthropomorphic elements of conversational AI, including visual, linguistic, and non-verbal aspects, influence learners' immersion, satisfaction, and intention to continue. Through Focus Group Interview (FGI), the researchers established a visual stage for the AI where learners can naturally engage in conversation, which was then integrated into the prototype. The prototype experiment involved interactions with AIs at varying levels of anthropomorphism: non-anthropomorphic, weak anthropomorphism, and strong anthropomorphism, while participants performed tasks involving art and social-themed conversation scenarios. The key evaluation metrics used in the study were Perceived Ease of Use, Perceived Usefulness, Perceived Interactivity, Perceived Intelligence, Perceived Trust, and Intention to Use, with a focus on participants' immersion, learning outcomes, and user satisfaction.

Results: The experimental results showed that the prototype incorporating interactions with a conversational AI featuring strong anthropomorphism scored the highest across all six metrics: Perceived Ease of Use, Perceived Usefulness, Perceived Interactivity, Perceived Intelligence, Perceived Trust, and Intention to Use. This suggests that anthropomorphizing conversational AI, encompassing visual, linguistic, and non-verbal elements, can have a positive impact on foreign language conversation learning.

Conclusion: This study provides evidence that anthropomorphizing the user interface is a crucial factor in enhancing conversation immersion and fostering continued usage intention in the development of foreign language conversation learning services utilizing conversational AI. These findings suggest that in the future design and development of conversational AI, strategic integration of anthropomorphic elements into the user interface should be pursued.

Application: In designing foreign language conversation learning using conversational AI, the user interface can be constructed by considering anthropomorphic elements such as those proposed in this study, including visual, linguistic, and non-verbal aspects of anthropomorphism.

Keywords: Anthropomorphism, Conversational AI, FGI, Foreign language speaking

1. Introduction

1.1 Background

대화형 AI (Conversational Artificial Intelligence)는 음성 또는 텍스트 기반의 대화 및 상호작용을 시뮬레이션하고 자동화할 수 있는 기능을 갖춘 인공지능(Artificial Intelligence, AI)으로 의료, 고객 관리, 전자 상거래 및 교육과 같은 다양한 분야에서 활용되고 있다(Kulkarni et al., 2019). 대화형 AI는 자연어 처리 시스템(Natural Language Processing, NLP)과 사용자 인터랙션 인터페이스(User Interaction Interface)으로 구성되어 있으며, 사용자에게 전달받은 데이터를 대화 관리 시스템(Dialogue Management System)을 통하여 처리하고, 이를 기반으로 사용자가 요구하는 정보를 전달한다(Kulkarni et al., 2019). 따라서 대화형 AI의 사용자 인터페이스는 AI 기술 자체에서 발생할 수 있는 사용자 입력에 따른 변동성이나 개인화로 인한 잘못된 결과 등과 같이 예측할 수 없는 오류 등에 의해 사용자가 이탈하지 않고 신뢰할 수 있도록 구성되어야 한다(Amershi et al., 2019). 대화형 AI가 등장하기 전에는, 외국어 회화 학습 분야에서 컴퓨터의 자동 음성 인식 시스템(Automatic Recognition System, ARS)이 회화 학습이나 발음 교정을 위한 기술로 활용되었다. ARS는 학습자에게 개인적인 학습 환경을 마련하여 개인화된 교육 서비스를 제공할 수 있지만, 동시에 언어 학습에 있어서 중요한 요소인 피드백 측면에서 사용자에게 부적절한 피드백을 제공할 수 있으며, 이는 교육의 기본적인 요건들을 충족시키지 못한다는 비판이 있었다(Neri et al., 2002). NLP 기반의 대화형 AI는 교정적 피드백(Corrective feedback)과 의미 협상(Negotiation of meaning)을 제공하여 사용자와 즉각적인 상호작용이 가능하기 때문에 언어 학습에 긍정적인 영향을 미칠 수 있고, 기존 ARS의 한계를 극복할 수 있을 것으로 기대된다(Bibauw et al., 2022). 또한 대화형 AI는 원어민과 실제로 대화하기 어려운 상황에 처한 사용자가 언제 어디서든 사용할 수 있기 때문에, 외국어 학습의 회화 파트너로 촉망받고 있다(Belda-Medina and Calvo-Ferrer, 2022). 외국어로 의사소통을 할 때, 대화형 AI는 사용자가 자신의 부족한 회화 실력 때문에 느낄 수 있는 걱정과 두려움과 같은 부정적인 감정을 극복하고 학습을 지속하는 대안이 될 수 있다(Ayedoun et al., 2019). 이에 따라 외국어 회화 학습과 관련하여 대화형 AI에 대한 연구가 활발하게 일어나고 있다. Dillon (2018)은 아마존의 음성 어시스턴트인 Alexa를 활용하여 1, 2학년 학생들에게 교과서 연습 문제를 학습하도록 하였다. 학습 과정에서 음성 인식 오류가 발생하는 등의 문제가 일부 발생하였으나, 학생이 대화형 AI의 피드백 시스템에 익숙해지면 학습이 원활히 진행된다는 결과가 나타났다. Ruan et al. (2021)는 56명의 중국 대학생을 대상으로 적응형 피드백(Adaptative feedback)을 제공하는 Englishbot과 기존의 듣고 따라 하는 인터페이스를 비교한 뒤 IELTS (International English Language Testing System) 채점 기준으로 학생들의 영어 실력 향상을 평가했을 때 EnglishBot을 사용하면 더욱 향상된다는 사실을 보였다. Yang et al. (2022)는 한국의 영어 학습자들이 과제 중심 음성 챗봇을 활용하여 말하기 과제를 수행하도록 하였고, 챗봇을 활용한 학습자들은 높은 과제 성공률을 보였다. Jeon (2024)는 초등 영어 교실에서 36명의 초등학생 학습자를 대상으로 Google의 Dialogflow를 사용하여 생성된 맞춤형 챗봇을 활용하게 하여 영어 학습 동기와 심리적 상태 등에 대한 질적 분석을 시행했다. 분석 결과, 학습자가 자신의 언어 역량으로 챗봇의 과제를 해결할 수 있

다고 인식하거나 챗봇을 자율적으로 제어할 수 있는 상황에서 학습 동기가 더욱 향상되고 학습에 더욱 적극적으로 참여하는 것으로 나타났다.

대화형 AI와의 외국어 회화 학습에 대한 다양한 연구가 이루어지고 있는 반면, 대화형 AI를 활용하여 학습을 하였을 때 새로운 기술에 대한 단기적인 흥미만 느낄 뿐 지속적인 학습을 위한 사용으로 이어지기 어렵다는 문제가 제기되기도 하였다. Fryer et al. (2017)은 학습용 챗봇이 지속적인 외국어 학습을 유도하는지 확인하기 위해, 실제 사람과의 회화 학습과 비교하였으나 학습용 챗봇과 함께 회화 학습을 하는 학습자들의 흥미는 처음 회화 작업 이후 유의미하게 감소했으며, 동일한 작업을 인간 파트너와 수행하는 경우 흥미는 동일한 수준을 유지하였다. 이는 학생들은 학습용 챗봇과의 학습을 진정성 없는 말하기 경험으로 인식하였고 결과적으로 학습용 챗봇과의 상호작용을 더 낮은 학습 경험으로 인식하였다고 해석된다(Fryer et al., 2017). 한편, 의인화(Anthropomorphism)는 본래 인간과 유사한 외모, 감정, 성격, 행동 등 인간과 유사한 특징에 의해 오는 지각 등 다양한 관점에서 연구자들에 의해 정의되었다(Li and Suh, 2022). 대화형 AI에게 인간의 특성을 부여하는 의인화가 중요한 이유는, 사용자가 대화형 AI를 사용하고자 하는 지속적인 사용 의사에 긍정적으로 직접적으로 영향을 줄 수 있기 때문이다(Li and Suh, 2022). 기존 연구에서는 사용자가 챗봇이 지적인 능력을 가지고 있다고 인식할수록 친밀감과 공존감과 증가하며, 이는 사용자의 사용 의도에 대한 긍정적인 영향을 미칠 수 있음이 나타났다(Lee et al., 2020). 또한 이러한 의인화는 단순히 친밀감뿐 아니라 사용자와의 관계 구축에도 영향을 미칠 수 있으며 이러한 경험은 사용자가 더 '인간적'이라고 느껴지는 대상과 더 많은 시간을 보내고 싶게 만들 수 있다(Qiu et al., 2020). 이처럼 사용자가 의인화에 따라 사용 의사가 증가하는 것은, 대화형 AI가 가진 인간과 유사한 상호작용 역량과 사용자가 느끼는 신뢰감이 사용자의 참여를 촉진하는 데 중요한 역할을 한다는 관점으로도 볼 수 있을 것이다(Chandra et al., 2022). 따라서 대화형 AI의 의인화는 사용자와의 상호작용을 이끌고, 사용자와의 대화에 있어서 관계 구축에 영향을 미칠 수 있기 때문에, 대화형 AI의 인터페이스를 설계하는 것에 있어서 중요한 요인으로 여겨진다.

본 연구는 외국어 회화 학습에 있어서 단기적인 흥미에 의해 지속적인 학습이 어렵다는 문제점을 대화형 AI 튜터의 의인화를 통해 개선할 수 있는지 알아보고자 한다. 외국어 회화 학습에서의 대화형 AI 튜터를 의인화한다면, 실제 사람과의 대화보다 낮은 몰입도, 단기적인 흥미와 같은 대화형 AI와의 회화 학습에서 나타날 수 있는 문제점들을 해소할 수 있을 것이라고 기대한다. 본 연구에서는 회화 학습에 있어서 대화형 AI의 의인화가 실제로 학습에 있어서 사용자들의 지속적인 사용 의사를 유도하며, 사용성 및 상호작용을 개선할 수 있는지 확인하고 나아가 의인화가 적용된 시각적 인터페이스를 제안하는 것이 목적이다. 이에 대화형 AI 회화 학습 경험 개선을 위해 선행연구에서 의인화 요소를 도출하였으며, 대화형 AI에 익숙한 20대 사용자들과의 FGI (Focus Group Interview, FGI)를 통해 대화형 AI 튜터의 시각적 특징 및 회화 시나리오를 설정하고 어플리케이션 프로토타입을 제작하였다. 대화형 AI 튜터에 대한 사용성 지표에 대한 선행연구를 통해, 대화형 AI 튜터와의 회화 학습에서 유효한 평가 지표 6가지를 선정하여, 비의인화, 약한 의인화, 강한 의인화로 구분된 3가지 의인화 수준에 따른 사용자 경험을 평가하였다.

1.2 Anthropomorphization

의인화의 특성은 크게 외형과 상호작용이라는 두 관점으로 구분될 수 있다(Choi and Kim, 2008). 외형에 대한 연구로 Seeger et al. (2021)은 대화형 AI의 인격 요소 존재 여부의 영향을 평가하고자 대화 상황에서 인간 정체성을 느끼게 하는 이미지로 'Laura'와 같은 인간적 이름 또는 사람의 프로필 이미지를 부여했다. 그 결과 대화형 AI의 기술적 특징인 자연스러운 상호작용성과 맞물려 긍정적인 상호작용 경험과 인간적 인식을 증가시킬 수 있음을 보였다. Garau et al. (2003)은 사실적인 시선 처리를 보여주는 인간형 아바타가 무작위 시선 시스템을 갖춘 아바타보다 더 높은 사회적 존재감을 유도한다는 것을 보여주었다. Strohmann et al. (2023)은 대화형 AI의 대인관계적으로 풍부한 대화를 가능하게 하는 요소를 파악하기 위해 AI 대화 서비스 'Replika'를 분석했다. 대화형 AI를 3D 모델 형태의 인간 아바타를 제공해, 눈 색깔, 헤어스타일 같은 세부 요소를 변경할 수 있게끔 하여 사용자 맞춤형 서비스를 제공했다. 대화하는 공간 속 아바타가 맥락에 맞는 표정을 지어가면서 사용자와의 상호작용이 가능할 때 실제와 가상의 삶이 동일시 되는 것 같은 느낌을 주었다(Strohmann et al., 2023).

의인화의 상호작용 측면에서는 언어적 표현과 비언어적 표현으로 구별할 수 있다(Seeger et al., 2021). 언어적으로는 대화형 AI가 1인칭 대명사(예: "나")와 감정표현(예: "오 미안해요")을 사용하거나 비언어적으로 이모티콘을 쓰고 텍스트를 쓰느라 지연되는 상황을 연출했다. Ho et al. (2018)는 대화형 AI가 감정적 자아를 드러낼 때 인간과 비슷한 정서적, 관계적, 심리적 이점이 발생한다는 사실을 관

찰했다. 이모티콘 사용이 대인 관계 매력도에 미치는 영향에 대한 최근 연구에서 챗봇은 이모티콘을 사용하여 동일한 메시지 콘텐츠를 전달할 때 인간 대화 상대만큼 사회적으로 매력적이고 신뢰할 수 있는 것으로 나타났다(Beattie et al., 2020). Haugeland et al. (2022)은 챗봇과의 대화 유형 변화가 사용자들에게 미치는 영향을 찾고자 신용카드를 주문하는 고도의 목표지향적인 대화와 연금 저축의 조언을 얻기 위한 주제 주도형 대화 시나리오를 실험했다. 그 결과 주제 주도형 대화를 제공하는 챗봇이 사용자에게 더 높은 지각된 의인화를 제공함을 발견하였다. Murano et al. (2011)는 기존의 대화형 인터페이스에서 음성이 포함된 효과를 이용했을 때 더 나은 사용자 만족도를 이끌 수 있는 것으로 확인되었다. 이처럼 대화형 AI가 의인화되었을 때 사용자에게 미치는 요인에 대한 다양한 연구가 이루어졌다.

외국어 학습에 있어서도 대화형 AI의 의인화에 대한 연구들이 이루어졌다. Wang et al. (2022)은 대화형 AI를 웃는 얼굴을 가진 여성 교사로 의인화하였고 외국어 학습에 대한 개인 맞춤형 피드백을 제공했다. 이에 대한 결과로 학습자는 대화형 AI와의 상호작용을 통해 더 높은 사회적, 인지적 존재감을 인식하며 대화형 AI의 외형에 애정을 가진 학습자가 더 자주 대화형 AI를 사용하고 높은 학습 성과를 달성한다는 결과가 나타났다. 그러나 대화형 AI의 의인화를 실험자가 임의로 제시하여, 이에 대한 사용자가 선호하는 의인화 형태나 유형 같은 구체적인 시각적 인터페이스를 알기 어려웠다. 또한 외국어 회화 학습에 있어서, 의인화와 관련된 언어적, 비언어적 표현 정도에 따른 지속적 사용 의사나 사용성에 대한 연구가 부족했다. 따라서 사용자의 대화형 AI에 대한 인식 및 수준을 바탕으로 어느 수준 및 어느 유형의 의인화 이미지를 선호하는지에 대한 논의가 필요하다.

2. Method

본 연구에서는 먼저 5명의 참여자를 대상으로 FGI를 시행하여 사용자 선호, 의인화 관련 인식을 조사했으며, 이를 기반으로 의인화 수준에 따른 3가지의 외국어 학습을 위한 대화형 AI 튜터의 프로토타입을 설계했다. 이후 프로토타입의 의인화 수준에 따른 사용 지속 의사, 사용성 등을 평가하기 위해 20명의 참여자를 대상으로 실험 연구를 진행하였다.

2.1 Focus group interview

2.1.1 Objective

FGI는 사회자의 안내에 따라 특정 주제에 대해 참가자들이 함께 토론을 하며 생각과 의견을 알아보는 방식이다(Krueger, 1988). 대화형 AI 튜터의 인터페이스에서 표현되는 의인화 이미지가 사용자와의 상호작용에 미치는 영향을 알아보기 위해 FGI를 진행했다. 본 연구는 대화형 AI 튜터와 영어로 대화할 때 어떤 인물과 외형 이미지를 선호하는 지와 평소 대화형 AI를 활용한 외국어 회화 학습에 대한 인식에 대해 질문했다.

2.1.2 Experimental procedure

FGI는 오프라인 환경에서 동시에 이루어졌으며 대형 모니터로 참가자 모두가 대화형 AI 튜터의 의인화 단계 및 캐릭터 속성에 따라 구별된 시각 자료를 볼 수 있도록 했다. 총 5명의 참가자가 모여 진행자의 질문에 정해진 순서나 형식 없이 대답하는 자유토론 방식으로 진행했다.

2.1.3 Participants

본 실험에는 5명의 대학생이 참가하였다. 참가자 나이는 평균 21.8세로 구성되었으며, 남성 4명, 여성 1명이었다. (Table 1)는 FGI 참가자 정보이다. 참가자 선발 과정에서 다양한 학과에 속한 학생들로 선발하였으며, 참여자들은 모두 AI 회화 학습 또는 모바일 회화 서비스를 이용해본 경험이 있었다. 참가자들은 인터뷰에 참가하기 전 자신의 영어 회화 수준에 대해서 스스로 평가했다.

Table 1. Participants of focus group interview

Subject	Participant A	Participant B	Participant C	Participant D	Participant E
Gender	Female	Male	Male	Male	Male
Age	22	22	22	22	21
Job	University student	University student	University student	University student	University student
English conversation level	High intermediate level	High intermediate level	Intermediate level	Intermediate level	High beginner level
Experience of using a conversational AI service	Yes	Yes	Yes	Yes	Yes
Experience of using a conversational AI service for learning English	Yes	No	Yes	No	No

2.1.4 Process

진행은 세 가지 프로세스로 2시간에 걸쳐 진행되었다. 1차로 대화형 AI 서비스에 대한 인식과 회화 연습 용도의 사용 경험을 물어 대화형 AI에 대한 개인적 경험을 환기했다. 2차로 대화형 AI 튜터의 의인화 단계를 보며 자유롭게 느낌을 얘기하도록 했다. 3차는 다양한 성격과 직종이 드러나는 캐릭터를 나열하여 어떤 속성의 인물과 지속적으로 회화하고 싶은 지 물었다.

(1) 대화형 AI 서비스를 통한 회화연습 인식 및 사용 경험
대화형 AI 서비스와 대화를 해본 경험을 물었다. 평소 대화형 AI와 대화하는 것에 어떤 느낌을 가지고 있었는지, 대화를 하면서 느껴진 또한 인간적인 요소와 기계적인 요소에 대한 의견을 나누어 얘기하도록 했다. 또한 대화형 AI 튜터와 영어로 대화할 시에 회화 연습 용도로써 개인에게 어떤 효과를 줄 수 있을지 의견을 물어보았다.

(2) 대화형 AI 튜터의 의인화 외형에 대한 선호도
의인화 외형 척도 자료는 Mohd Tuah et al. (2016)의 의인화 분류를 참고해 제작했다. Mohd Tuah et al. (2016)의 연구에서는 실제 인간과 같은 외형에서부터 캐릭터와 같은 외형 사이를 다양한 이미지로 나열하여 의인화 외형 척도를 제작하였다. 이를 참고해 본 연구에서는 참여자들에게 인간 외형과의 높은 유사성과 낮은 유사성 사이에서 구별될 수 있는 이미지를 6단계로 나누어 대화형 AI 튜터 외형의 의인화 척도를 구성하였다. 직접 AI와 대화하는 상황을 가정할 수 있도록 해당 이미지의 프로필을 가진 AI가 메시지를 보낸 상황을 연출하여 실사적 이미지(Figure 1의 좌측에 해당)와 기계적 이미지(Figure 1의 우측에 해당) 중 어느 축을 더 선호하는지에 대해 자유토론 하였다.

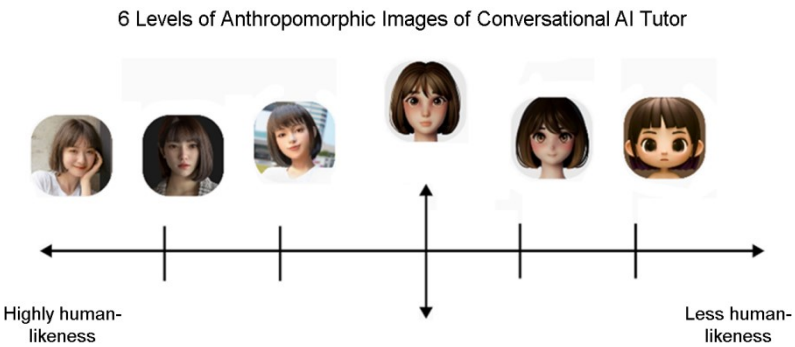

Figure 1. 6 levels of anthropomorphic images of conversational AI tutor

(3) 캐릭터 속성에 의한 지속 여부 파악
나이, 성격이 드러나는 이미지를 수집하였다. 해당 이미지 자료를 FGI 참가자가 쉽게 선택할 수 있도록 4가지 특성으로(전문적인, 친근한, 직업이 뚜렷한, 취미가 뚜렷한) 나누어 대화형 AI 튜터의 인구통계학적 정보를 선택할 수 있는 자료를 제작했다(Figure 2). FGI 참여자들에게 이미지 자료를 제시하여 회화 연습 상황에서는 어떤 속성의 인물이 선호도가 높은 지 질문했다. 또한, 유아, 청소년, 청년, 중년, 장년 등의 명확하게 구별되는 이미지를 나열하여 어떤 나이대의 사람이 적당한가 질문했다(Figure 3).

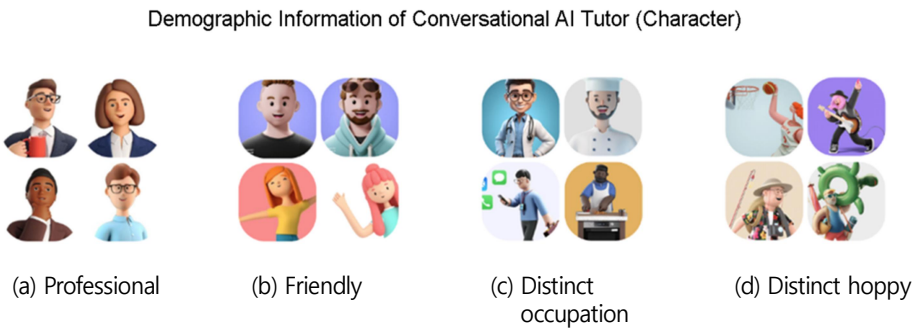

Figure 2. Demographic information of conversational AI tutor (Character)

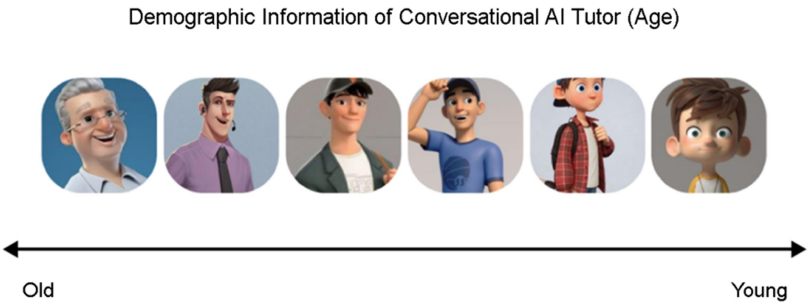

Figure 3. Demographic information of conversational AI tutor (Age)

2.1.5 Analysis & Results

(Table 2)는 FGI에 대한 분석 결과이다. FGI를 통해 대화형 AI를 통한 학습에 대한 인식과 대화형 AI 튜터의 의인화와 관련한 선호도 및 학습 지속 의사를 파악하고자 했다.

(1) 대화형 AI 서비스를 통한 회화 연습 인식 및 사용 경험 결과
대화형 AI 튜터의 목소리와 이미지의 조화가 중요하며, 따라서 회화하는 것을 인지하기 위한 최소한의 의인화 그래픽을 선호하였다. 역할의 다양성에 대한 니즈가 있었으며 회화 학습 중 사용자의 사회적 존재감이 느껴지는 것을 원했다. 성취감이나 스토리텔링과 같은 부가적인 요소에 대한 선호 또한 있었다.

(2) 대화형 AI 튜터의 의인화 단계 선호도 결과
(Figure 1)의 Highly human likeness에 근접한 어설픈 인격화보다는 확실한 캐릭터성이 나타나는 Less human likeness 쪽을 선호했다. 강한 캐릭터성이 게임을 하는 듯한 상황을 연상시켜, 오히려 캐릭터와 상호작용하는 듯한 몰입감을 선호하는 것으로 답했다. 4~5단계 (Highly human likeness에 근접한)에 위치한 이미지는 오히려 몰입을 방해하며, 실제 인간과 대화하는 것이 아님에도 불구하고, 인간과

닮은 캐릭터와 대화하는 것이 혼란을 야기된다고 참여자들은 답했다.

(3) 대화형 AI 튜터의 인구통계학적 속성에 의한 학습 지속 의사

(Figure 2)에서 'Job'과 'Professional' 축에 속하는 특정 분야에 전문성이 있는 이미지를 선호하지 않았다. 회화 연습을 돕는 대화형 AI 튜터는 전문적으로 가르침을 줄 것 같은 이미지를 강조하는 것보다 더 자연스럽게 대화할 수 있는 'Hobby'와 'Friendly' 축에 속하고 또래인 20대의 이미지를 선호하는 것으로 나타났다. 그러나 취미가 뚜렷한 속성은 사용자 개인의 취미에 영향을 주어 선호도에 있어서는 뚜렷한 영향을 미치지 않는 것으로 보였다.

Table 2. Analysis of FGI

Question	Answer
1) Perceptions and usage experiences of speaking a foreign language through interactive AI services.	1. Preference for a combination of voice and text due to location constraints: It would be good to have text as well as voice recognition ("Text can be used anywhere"). 2. Reflection of recent trends and personal interests: It would be good to reflect recent trends as if talking with foreign friends ("DMs with actual slang and memes used by foreigners"). 3. Communication on desired topics: I want to communicate on topics of my interest ("It's difficult because there's no common topic when talking with foreign friends"). 4. Sense of achievement: There needs to be elements that can give users a sense of achievement ("Giving rewards for speaking a lot in English, etc").
2) Perception of anthropomorphism in AI	1. Importance of voice: The realism of the voice is importan ("How similar the voice is to a real human voice matters"). 2. Need for anthropomorphism in conversational learning: A certain degree of anthropomorphism is necessary when learning conversation ("At least minimal anthropomorphic graphics are needed to recognize that it is conversing"). 3. Uncanny realism causes discomfort: Having an image that looks too much like a real person, but isn't, can cause discomfort ("It's weird because it looks real but it isn't"). 4. Preference for clearly characterized figures: Appreciating the mechanical characteristics of AI and preferring a character that embodies these traits feels good ("It seems best when it is closest to a character").
3) Willingness to continue learning based on the demographic attributes of conversational AI tutors	1. Emphasizing the user's social presence: The user's social presence should be evident ("There should be a sense of presence among a variety of AI personalities"). 2. Storytelling elements: Storytelling in the process of conversing and forming relationships is necessary ("An additional environment conducive to forming intimate relationships is needed"). 3. Diverse character and conversational situations: A variety of characters and conversational situations are necessary ("It would be good if there were a variety of people and roles"). 4. Preference for Peer Tutors Over Professional Tutors: There is a tendency to prefer peer tutors, who can converse naturally and smoothly, over professional tutors who emphasize expertise in conversational learning. Additionally, tutors of the same age group are preferred because they are seen as more likely to share common conversation topics, unlike tutors who are significantly older or younger.

2.2 Procedure

2.2.1 Prototyping

실험을 위해 외국어 회화 학습 서비스의 대화형 AI 튜터 프로토타입을 제작했다. 프로토타입의 유저 플로우(User Flow)는 (Figure 4)와 같다. 프로토타입을 시각화 이전에 참여자가 서비스에 접속하고 대화형 AI 튜터와 대화를 진행하고 마무리하는 전체적 흐름의 자연스

러움을 확인하고자 유저 플로우를 제작했다(Interaction Design Foundation, 2016). 참여자는 메신저 화면에 들어가 대화형 AI 튜터에게 메시지를 받고 녹음 버튼을 눌러 대화형 AI 튜터의 질문에 답변하게 되며, 대화형 AI 튜터는 이에 대한 인터랙션을 제공한다. 해당 과정에서 참여자는 실험에서 정의한 3가지 의인화 수준의 대화형 AI 튜터와 상호작용한다.

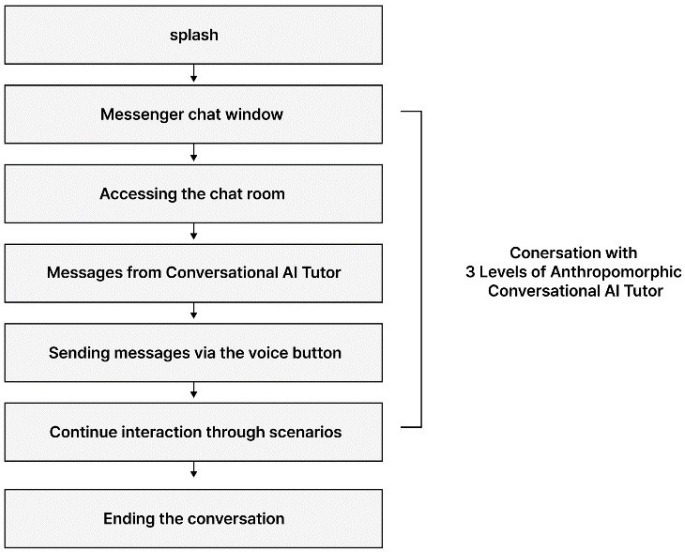

Figure 4. User flow of prototype

Mohd Tuah et al. (2016)과 Van Pinxteren et al. (2020)의 연구를 참고하여 '의인화 요소'(프로필 이미지, 언어적 표현, 비언어적 표현)를 정의 하였으며, FGI 결과를 토대로 AI 캐릭터 이미지 및 사용자가 AI와의 영어 회화에서 사용자가 자연스럽게 대화할 수 있는 상황(친근한 이모지 반응, 밈 사용 등)의 디자인 요소를 결정했다. 이에 따라 본 연구는 3가지 단계의 의인화 수준을 정의하였으며, 세부 요소는 (Table 3)와 같다.

Table 3. 3 levels of anthropomorphic conversational AI tutor

3 levels of anthropomorphism (Independent variables)				
Elements of anthropomorphism		Level 1: non-anthropomorphic conversational AI tutor	Level 2: slightly anthropomorphic conversational AI tutor	Level 3: highly anthropomorphic conversational AI tutor
	Verbal behaviors	- Natural language generation - Dialogue management - Multimodal communication (text, voice, image)		
	Appearance	- Conversational user interface - Specifying the name as 'Tutor'	- Conversational user interface - Profile image - Specifying the name as 'Amy'	- Conversational user interface - Profile image - Specifying the name as 'Amy'
	Non-Verbal behaviors	None	None	- Expressing emotions - Sending meme pictures - Reacting with emoticons - Interacting while typing messages

(1) 비의인화 단계

비의인화 단계에서는 기본적 대화 상황만 나타냈으며 대화하는 대상이 특정 인격체가 아닌 튜터로 표현했고, 텍스트 기반의 대화만 인터페이스만을 허용하였다. (Figure 5)는 비의인화 단계의 인터페이스 프로토타입이다.

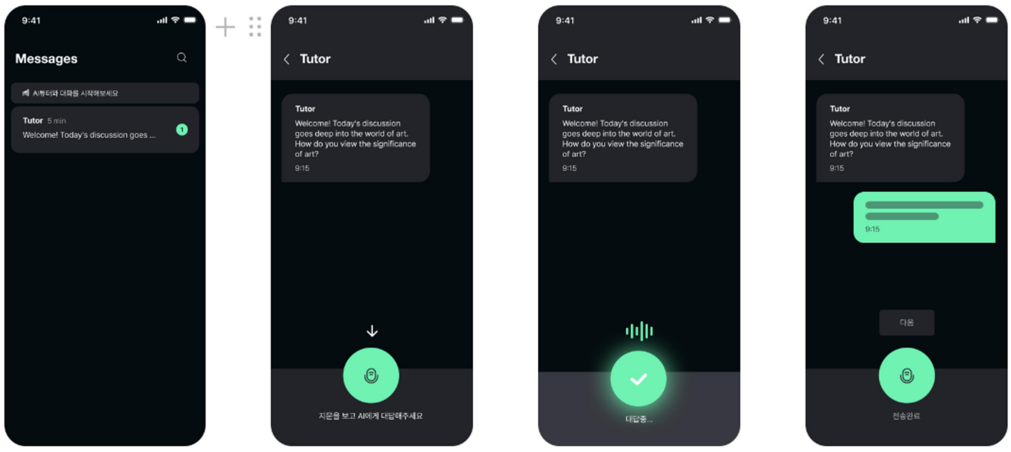

Figure 5. Prototype of non-anthropomorphic conversational AI tutor

(2) 약한 의인화 단계

약한 의인화 단계에서는 의인화 요소 중 '외형'에 관한 부분을 추가했다. FGI에서 검증한 캐릭터 이미지를 프로필로 보여주었으며, 대화형 AI 튜터에게 'Amy'라는 이름을 지정했다. (Figure 6)는 약한 의인화 단계의 인터페이스 프로토타입이다.

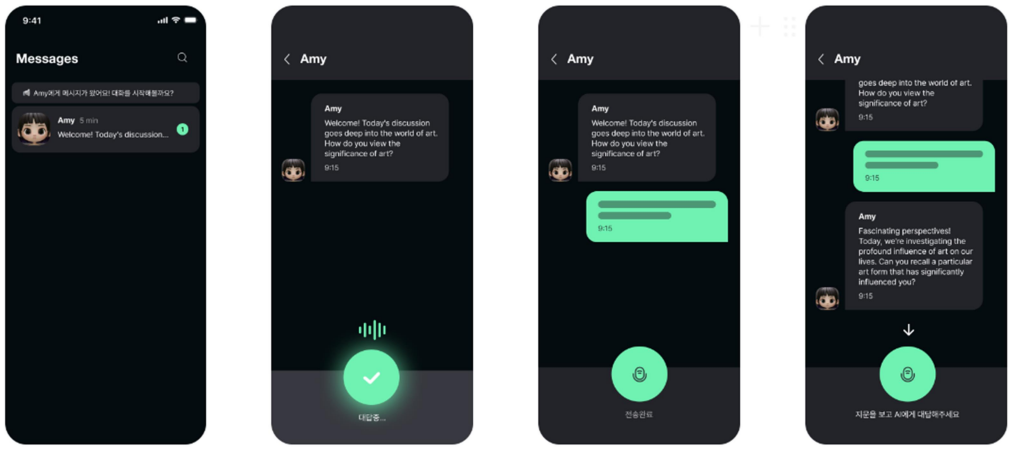

Figure 6. Slightly anthropomorphic conversational AI tutor

(3) 강한 의인화 단계

강한 의인화 단계에서는 앞 조건과 동일한 상태에서 언어적, 비언어적 표현을 추가했다. 이모티콘을 사용하거나 유머러스한 밈으로 대화에 답하는 듯한 연출을 하여 대화형 AI 튜터로부터 다양한 감정이 느껴지도록 했다. (Figure 7)는 강한 의인화 단계의 인터페이스 프로토타입이다.

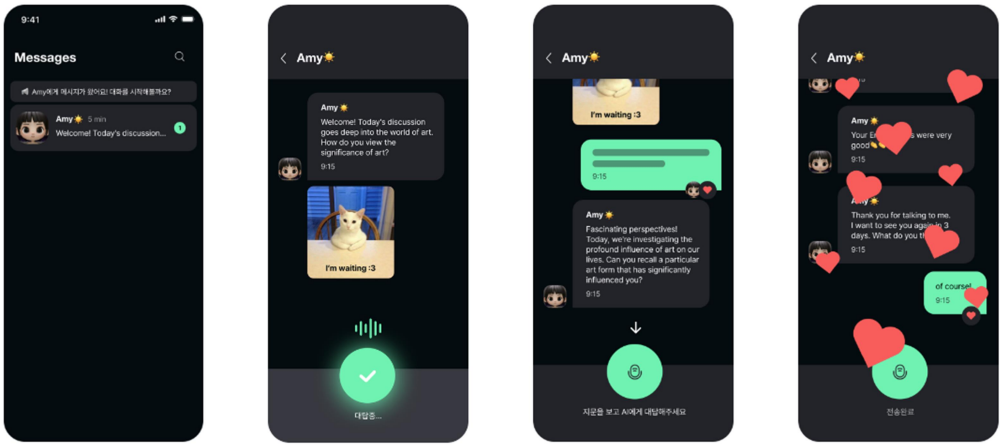

Figure 7. Highly anthropomorphic conversational AI tutor

2.2.2 Experimental design

(Figure 8)은 실험에 참여하고 있는 피실험자의 모습이며, (Figure 9)은 실험에 대한 절차이다. 본 실험에서 피실험자는 비의인화, 강한 의인화, 약한 의인화로 설계된 대화형 AI 튜터와 약 5분 간 대화하였다. 각 수준의 대화형 AI 튜터와 사회, 문화 2가지 주제에 대해 순차적으로 대화하였으며, 각 주제마다 대화형 AI 튜터는 피실험자에게 3가지 질문을 하였다. 문화 관련 주제에서는 예술에 관한 피실험자의 생각을 묻는 간단한 질문으로 구성하였으며, 사회 관련 주제에서는 인종 차별에 관한 주제로 질문을 구성하였다. (Table 4)는 각 문화, 사회 관련 주제에 관해 피실험자에게 주어진 질문이다. 피실험자에게 AI 튜터가 제시하는 질문에 대한 3가지 답안을 보여주고, 피실험자가 답안 중 자유롭게 선택하여 대화형 AI 튜터와 외국어로 대화할 수 있도록 유도하였다.

Figure 8. Image of experiment

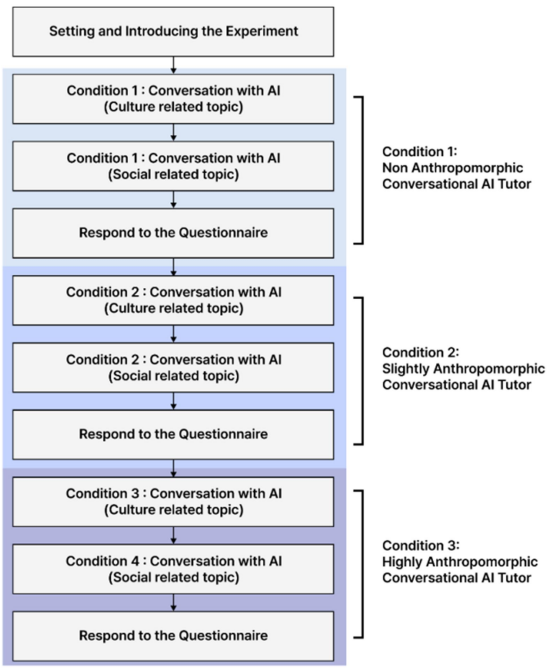

Figure 9. Procedure of experiment

Table 4. Questions generated by Chat GPT

Topic	Questions generated by Chat GPT
Cultural related topics	1. How do you view the significance of art?
	2. Can you recall a particular art form that has significantly influenced you?
	3. I'm curious to explore further how your chosen art form has left an impact on you.
Social related topics	1. How do you perceive the impact of racial discrimination on society?
	2. Can you recall a specific experience related to racial discrimination that has deeply affected you?
	3. I'm really interested in understanding how these experiences have shaped your views and actions in addressing with racial discrimination.

2.2.3 Experiment apparatus

본 실험에서는 Chat GPT, Figma와 Voice Conversation kit를 활용하여 프로토타입을 제작하였다. Figma를 통해 프로토타입의 인터페이스를 제작하였고, Chat GPT를 활용하여, 문화, 사회 각각의 주제에 해당하는 스크립트를 제작하였다. 스크립트 내용을 Voice Conversation Kit를 활용하여 음성으로 변환하였다. Figma의 프로토타입 기능을 활용하여, 실제 대화형 AI 튜터와 상호작용하는 서비스를 설계했다.

2.2.4 Participants

참여자는 총 20명으로 참여자 모두 대화형 AI 서비스에 대해 경험해본 참여자로 선정하였다. 실험 이전의 영어 회화 수준을 스스로 평가하였으며, 원어민 수준이나 외국어 회화 경험이 없는 참여자는 배제했다. 남성 10명, 여성 10명으로 구성되었으며, 평균 나이는 22.8세였다.

2.2.5 Questionnaire

본 논문에서는 회화 학습에서의 대화형 AI 튜터의 의인화 정도가 사용자에게 미치는 영향을 밝히기 위해 기존 챗봇 및 대화형 AI 등에 대한 연구를 토대로 Perceived ease of use, Perceived Usefulness, Perceived Interactivity, Perceived Intelligence, Perceived Trust, Intention to Use 로 구성된 6가지의 지표를 설정하였다. Perceived ease of use는 사용자가 사용에 대해 노력이 필요하지 않을 것으로 믿는 것을 나타내며, 이러한 사용자의 인식은 사용 행동에 대한 영향을 미칠 수 있다(Davis, 1989). Perceived Usefulness는 사용자가 특정 시스템을 사용함으로써 작업이나 업무의 성과를 향상시킬 것으로 인식하는 정도를 나타낸다(Davis, 1989). 본 논문에서는 Davis (1989)가 제시한 Perceived Ease of Use와 Perceived Usefulness 문항을 본 연구의 목적에 맞게 수정하였다. Perceived ease of Use는 사용자가 대화형 AI와 대화, 상호작용하는 것에 있어서 어려움 없이 수행하였는지, Perceived Usefulness는 사용자가 대화형 AI 튜터와 소통하는 것에 있어서 학습 및 실력 향상에 도움이 되었는지에 대한 내용으로 구성하였다.

Perceived Intelligence는 사용자가 의인화된 대화형 AI의 자연어를 생성하고 처리하는 능력 및 에이전트의 행동을 효율적이고 유용하며 효과적인 결과를 가져온다고 지각하는 것을 의미한다(Moussawi and Koufaris, 2019). 즉, 언어와 인지능력과 관련된 특정한 능력에 의해 유능함을 지각하는 것이다(Yu, 2020). Moussawi et al. (2021)는 대화형 AI라는 기술을 사용자가 채택하기 위해 Perceived Intelligence가 중요한 요소임을 밝히기도 했다. 본 논문에서는 Moussawi and Koufaris (2019)가 제시한 문항을 수정하여, 대화형 AI 튜터로부터 판단되는 지적 능력과 대화 역량을 묻는 질문으로 재구성하였다.

Perceived Trust는 상호작용하는 사용자의 의존 및 신뢰감에 대한 감정을 나타낸다(Pillai and Sivathanu, 2020). De Graaf and Malle (2017)는 AI의 시스템 구성에 따라 사용자가 해당 시스템에 대한 신뢰감에 영향을 줄 수 있다고 했다. 본 논문에서는 Tarhini et al. (2017)와 Pillai et al. (2023)가 제시한 문항을 수정하여, 대화형 AI 튜터와 대화에서 느껴지는 신뢰성을 묻는 질문으로 재구성하였다.

Percevied Interactivity는 대상이 서로의 요구에 응답하거나 용이하게 해주기 위한 정도라고 정의할 수 있다(Ha and James, 1998). 본 논문에서는 Etemad-Sajadi (2016)가 제시한 문항을 수정하여, 대화형 AI 튜터와의 상호작용과 관련된 질문으로 재구성하였다.

시스템의 구성에 따라 사용자가 태도를 형성하고, 이에 따라 시스템에 대한 수용 여부, Intention to Use에 영향을 줄 수 있다(Pillai et al., 2024). Tarhini et al. (2017)에 따르면 교육에 있어 새로운 기술을 도입하기 위해서는 학습자의 Intention to Use를 조사하는 것이 중요하다. Gayashan and Samarasinghe (2024)는 사용자가 교육과 관련된 AI 챗봇을 사용함에 있어 Intention to Use에 미치는 영향을 파악하고자 하였다. Tarhini et al. (2017)와 Pillai et al. (2023)가 제시한 문항을 수정하여, 대화형 AI 튜터와의 지속적인 사용 의사를 묻는 질문으로 재구성하였다. 즉, Intention to Use는 사용자가 학습에 있어 대화형 AI 튜터라는 새로운 방식의 학습법과 기술을 받아들일 의사를 묻는 지표다. 본 연구에서는 리커트 7점 척도를 사용하였으며, 1점은 '전혀 그렇지 않다', 7점은 '매우 그렇다'를 나타냈다.

2.2.6 Hypotheses

본 연구에서는 본 연구에서는 언어적, 비언어적 표현이 포함된 강한 의인화 기반의 대화형 AI가 회화 학습에서 사용성을 개선하고 상호작용을 유도할 수 있는지를 확인하고자 가설을 세웠다.

기존 연구에서 실제로 다양한 의인화 요소로 인해 사용자의 경험을 개선하거나, 신뢰감을 얻은 사례가 있었다. Cheng et al. (2022)는 소비자가 상담원 역할을 하는 대화형 AI를 사용하는 것에 대해, 대화형 AI의 Perceived Warmth와 Perceived Competence과 같은 특성이 인간 상담원으로 바꿀 의사와 Perceived Trust에 영향을 미치는지 실험하였다. 이에 따라 Perceived Warmth와 Perceived Competence가 대화형 AI에 대한 Perceived Trust에 긍정적인 영향을 미치며, 상담원으로 전환하는 것에 부정적인 영향을 미친다는 결과가 나타났다. Adam et al. (2021)는 의인화에 기반한 언어적 표현과 사회적 존재감으로 인해 대화형 AI에 의해 사용자가 서비스 피드백 요청 준수 등 사용자의 순응에 미치는 영향을 검증하였다. Munnukka et al. (2022)는 대화형 AI의 의인화, 사회적 존재감, 대화 길이, 태도 등을 통해 나타나는 사용자와의 상호작용이 대화형 AI에 기반한 추천에 대한 인식을 개선하고, 신뢰감을 향상하는 것을 나타냈다. Zhou et al. (2023)는 대화형 AI의 의인화적 특성이 Perceived Usefulness에 긍정적으로 영향을 미치며, 이는 사용자의 Intention to use에 영향을 미친다고 나타냈다. 이러한 선행연구를 토대로 다음과 같은 가설을 수립했다.

1. 대화형 AI를 활용한 외국어 회화 학습에서 강한 의인화가 약한 의인화, 텍스트 기반 채팅형 방식보다 Perceived Ease of Use이 높게 나타날 것이다(H1).
2. 대화형 AI를 활용한 외국어 회화 학습에서 강한 의인화가 약한 의인화, 텍스트 기반 채팅형 방식보다 Perceived Usefulness이 높게 나타날 것이다(H2).
3. 대화형 AI를 활용한 외국어 회화 학습에서 강한 의인화가 약한 의인화, 텍스트 기반 채팅형 방식보다 Perceived Interactivity이 높게 나타날 것이다(H3).
4. 대화형 AI를 활용한 외국어 회화 학습에서 강한 의인화가 약한 의인화, 텍스트 기반 채팅형 방식보다 Perceived Intelligence이 높게 나타날 것이다(H4).
5. 대화형 AI를 활용한 외국어 회화 학습에서 강한 의인화가 약한 의인화, 텍스트 기반 채팅형 방식보다 Perceived Trust이 높게 나타날 것이다(H5).
6. 대화형 AI를 활용한 외국어 회화 학습에서 강한 의인화가 약한 의인화, 텍스트 기반 채팅형 방식보다 Intention to Use 이 높게 나타날 것이다(H6).

2.2.7 Analysis method

본 연구에서는 텍스트 기반, 약한 의인화, 강한 의인화라는 의인화 수준에 대하여, 사용자가 지속적인 학습에 있어 영향을 미칠 수 있는 6가지의 지표를 설정하고 각 수준별로 지표를 측정하여 점수의 차이를 밝히고자 하였다. 각 지표별로의 평균 차이가 통계적으로 유의한지 검증하기 위해 One-Way ANOVA를 시행하였다. 또한 3가지 의인화 수준의 점수 차이가 통계적으로 유의한지 밝히기 위해 Tukey's Post Hoc를 95% 신뢰수준으로 시행하였다.

3. Results

3.1 Perceived ease of use

Perceived Ease of Use 점수는 텍스트 기반의 대화형 AI (M=4.1, SD=1.31), 약한 의인화를 가진 대화형 AI (M=4.7, SD=0.90)보다 강한 의인화를 가진 대화형 AI (M=6.1, SD=0.57)에서 높은 Perceived Ease of Use 점수를 나타냈다. (Table 5)은 이에 대한 결과다.

Table 5. Mean ratings for anthropomorphism level on perceived ease of use, along with ANOVA results

Dependent variables	Experiment condition	Means	SD	*F*	*p*
Perceived ease of use	Condition 1 (Non anthropomorphic conversational AI tutor)	4.1	1.31	24.76	<.001
	Condition 2 (Slightly anthropomorphic conversational AI tutor)	4.7	0.90		
	Condition 3 (Highly anthropomorphic conversational AI tutor)	6.1	0.57		

Condition 사이의 Perceived Ease of Use 점수 간 통계적인 차이를 알아보기 위해 Tukey's test를 시행했고, (Table 6)는 Tukey's test 결과를 나타낸다.

Table 6. The result of Tukey HSD test between experiment condition on perceived ease of use

Independent variable (I)	Independent variable (J)	The average difference (I-J)	*p*
Condition 1	Condition 2	-0.64	.09
	Condition 3	-2.05	<.001
Condition 2	Condition 3	-1.41	<.001

Tukey's test 결과 약한 의인화와 텍스트 기반 방식 간의 Perceived Ease of Use 점수의 평균 차이는 0.64이며, 95% 신뢰구간에서 통계적으로 유의하지 않다. 반면 강한 의인화는 텍스트 기반 방식과 약한 의인화 간의 Perceived Ease of Use 간의 평균 차이가 각각 2.05, 1.41이며 통계적으로 유의하다. 따라서 Perceived Ease of Use에서는 약한 의인화는 텍스트 기반 방식 점수와 크게 차이 나지 않지만, 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 나타냈다. 즉, H1이 지지됨을 확인할 수 있었다.

3.2 Perceived usefulness

Perceived Ease of Use 점수는 텍스트 기반의 대화형 AI (M=4.0, SD=1.4), 약한 의인화를 가진 대화형 AI (M=4.4, SD=1.11)보다 강한 의인화를 가진 대화형 AI (M=5.8, SD=0.75)에서 더 높은 Perceived Usefulness 점수를 나타냈다. (Table 7)는 이에 대한 결과다.

Condition 사이의 Perceived Usefulness 점수 간 통계적인 차이를 알아보기 위해 Tukey's test를 시행했고, (Table 8)는 Tukey's test 결과를 나타낸다.

Table 7. Mean ratings for anthropomorphism level on perceived usefulness, along with ANOVA results

Dependent variables	Experiment condition	Means	SD	*F*	*p*
Perceived usefulness	Condition 1 (Non anthropomorphic conversational AI tutor)	4.0	1.43	15.13	<.001
	Condition 2 (Slightly anthropomorphic conversational AI tutor)	4.4	1.11		
	Condition 3 (Highly anthropomorphic conversational AI tutor)	5.8	0.75		

Table 8. The result of Tukey HSD test between experiment condition on perceived usefulness

Independent variable (I)	Independent variable (J)	The average difference (I-J)	*p*
Condition 1	Condition 2	-0.45	.426
	Condition 3	-1.89	<.001
Condition 2	Condition 3	-1.44	<.001

Tukey's test 결과 약한 의인화와 텍스트 기반 방식 간의 Perceived Usefulness 점수의 평균 차이는 0.45이며, 95% 신뢰구간에서 통계적으로 유의하지 않다. 반면 강한 의인화는 텍스트 기반 방식과 약한 의인화 간의 Perceived Ease of Use 점수 간의 평균 차이가 각각 1.89, 1.44이며 통계적으로 유의하다. 따라서 Perceived Usefulness에서는 약한 의인화는 텍스트 기반 방식 점수와 크게 차이 나지 않지만, 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 나타냈다. 이를 통해 H2이 지지됨을 확인할 수 있었다.

3.3 Perceived interactivity

Perceived Ease of Use 점수는 텍스트 기반의 대화형 AI (M=3.8, SD=1.42), 약한 의인화를 가진 대화형 AI (M=4.4, SD=1.07)보다 강한 의인화를 가진 대화형 AI (M=6.2, SD=0.71)에서 더 높은 Perceived Interactivity 점수를 나타냈다. (Table 9)는 이에 대한 결과다.

Table 9. Mean ratings for anthropomorphism level on perceived interactivity, along with ANOVA results

Dependent variables	Experiment condition	Means	SD	*F*	*p*
Perceived interactivity	Condition 1 (Non anthropomorphic conversational AI tutor)	3.8	1.42	24.99	<.001
	Condition 2 (Slightly anthropomorphic conversational AI tutor)	4.4	1.07		
	Condition 3 (Highly anthropomorphic conversational AI tutor)	6.2	0.71		

Condition 사이의 Perceived Interactivity 점수 간 통계적인 차이를 알아보기 위해 Tukey's test를 시행했고, (Table 10)는 Tukey's test 결과를 나타낸다.

Table 10. The result of Tukey HSD test between experiment condition on perceived interactivity

Independent variable (I)	Independent variable (J)	The average difference (I-J)	*p*
Condition 1	Condition 2	-0.57	.245
	Condition 3	-2.37	<.001
Condition 2	Condition 3	-1.80	<.001

Tukey's test 결과 약한 의인화와 텍스트 기반 방식 간의 Perceived Interactivity 점수의 평균 차이는 0.57이며, 95% 신뢰구간에서 통계적으로 유의하지 않다. 반면 강한 의인화는 텍스트 기반 방식과 약한 의인화 간의 Perceived Interactivity 점수 간의 평균 차이가 각각 2.37, 1.80이며 통계적으로 유의하다. 따라서 Perceived Interactivity에서는 약한 의인화는 텍스트 기반 방식 점수와 크게 차이 나지 않지만, 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 나타냈다. 즉, H3이 지지됨을 확인할 수 있었다.

3.4 Perceived intelligence

Perceived Intelligence 점수는 텍스트 기반의 대화형 AI (M=4.3, SD=1.40), 약한 의인화를 가진 대화형 A I(M=4.7, SD=0.91)보다 강한 의인화를 가진 대화형 AI (M=5.9, SD=0.99)에서 더 높은 Perceived Intelligence 점수를 나타냈다. (Table 11)는 이에 대한 결과를 나타낸다.

Table 11. Mean ratings for anthropomorphism level on perceived intelligence, along with ANOVA results

Dependent variables	Experiment condition	Means	SD	*F*	*p*
Perceived intelligence	Condition 1 (Conversational AI based text)	4.3	1.40	11.25	<.001
	Condition 2 (Conversational AI with weak anthropomorphism)	4.7	0.91		
	Condition 3 (Conversational AI with strong anthropomorphism)	5.9	0.99		

Condition 사이의 Perceived Interactivity 점수 간 통계적인 차이를 알아보기 위해 Tukey's test를 시행했고, (Table 12)는 Tukey's test 결과를 나타낸다.

Table 12. The result of Tukey HSD test between experiment condition on perceived intelligence

Independent variable (I)	Independent variable (J)	The average difference (I-J)	*p*
Condition 1	Condition 2	-0.42	.471
	Condition 3	-1.62	<.001
Condition 2	Condition 3	-1.20	<.001

Tukey's test 결과 약한 의인화와 텍스트 기반 방식 간의 Perceived Intelligence 점수의 평균 차이는 0.42이며, 95% 신뢰구간에서 통계적으로 유의하지 않다. 반면 강한 의인화는 텍스트 기반 방식과 약한 의인화 간의 Perceived Intelligence 점수 간의 평균 차이가 각각

1.62, 1.20이며 통계적으로 유의하다. 따라서 Perceived Intelligence에서는 약한 의인화는 텍스트 기반 방식 점수와 크게 차이 나지 않지만, 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 나타냈다. 즉, H4이 지지됨을 확인할 수 있었다.

3.5 Perceived trust

Perceived Trust 점수는 텍스트 기반의 대화형 AI (M=3.8, SD=1.43), 약한 의인화를 가진 대화형 AI (M=4.5, SD=1.10)보다 강한 의인화를 가진 대화형 AI (M=5.8, SD=0.79)에서 더 높은 Perceived Trust 점수를 나타냈다. (Table 13)는 이에 대한 결과다.

Table 13. Mean ratings for anthropomorphism level on perceived trust, along with ANOVA results

Dependent variables	Experiment condition	Means	SD	*F*	*p*
Perceived trust	Condition 1 (Conversational AI based text)	3.8	1.43	15.91	<.001
	Condition 2 (Conversational AI with weak anthropomorphism)	4.5	1.10		
	Condition 3 (Conversational AI with strong anthropomorphism)	5.8	0.79		

Condition 사이의 Perceived Trust 점수 간 통계적인 차이를 알아보기 위해 Tukey's test를 시행했고, (Table 14)는 Tukey's test 결과를 나타낸다.

Table 14. The result of Tukey HSD test between experiment condition on perceived trust

Independent variable (I)	Independent variable (J)	The average difference (I-J)	*p*
Condition 1	Condition 2	-0.73	.111
	Condition 3	-2.00	<.001
Condition 2	Condition 3	-1.27	<.001

Tukey's test 결과 약한 의인화와 텍스트 기반 방식 간의 Perceived Trust 점수의 평균 차이는 0.73이며, 95% 신뢰구간에서 통계적으로 유의하지 않다. 반면 강한 의인화는 텍스트 기반 방식과 약한 의인화 간의 Perceived Trust 점수 간의 평균 차이가 각각 2.00, 1.27이며 통계적으로 유의하다. 따라서 Perceived Trust에서는 약한 의인화는 텍스트 기반 방식 점수와 크게 차이 나지 않지만, 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 갖는 것을 알 수 있다. 즉, H5이 지지됨을 확인할 수 있었다.

3.6 Intention to use

Intention to Use는 텍스트 기반의 대화형 AI (M=3.8, SD=1.68), 약한 의인화를 가진 대화형 AI (M=4.7, SD=1.25)보다 강한 의인화를 가진 대화형 AI (M=6.2, SD=0.61)에서 더 높은 Intention to Use 점수를 나타냈다.

Tukey's test 결과 약한 의인화와 텍스트 기반 방식 간의 Intention to Use 점수의 평균 차이는 0.83이며, 95% 신뢰구간에서 통계적으로 유의하지 않다. 반면 강한 의인화는 텍스트 기반 방식과 약한 의인화 간의 Perceived Trust 점수 간의 평균 차이가 각각 2.35, 1.52이며,

95% 신뢰구간에서 통계적으로 유의하다. 따라서 Intention to Use에서는 약한 의인화는 텍스트 기반 방식 점수와 크게 차이 나지 않지만, 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 나타냈다. 즉, H6이 지지됨을 확인할 수 있었다.

4. Discussion

대화형 AI 튜터와의 의인화 수준을 평가하기 위해 설정한 모든 6가지의 평가 지표에서 강한 의인화는 텍스트 기반 방식과 약한 의인화와 비교하여 통계적으로 유의하게 더 높은 점수를 나타냈다. Perceived Ease of Use는 기존 대화형 AI 연구에서 자연스럽게 사용자의 입력을 유도하고, 또 사용자가 얻고자 하는 결과에 대한 행위를 단순화할 수 있는가에 대한 지표를 나타낸다(Brachten et al., 2021). 또한 Perceived Usefulness는 선행연구에서 사용자가 챗봇 기술을 얼마나 쉽게 사용하고 학습을 수행할 수 있는지를 확인하는 것이다(Chen et al., 2020). 즉, 해당 지표를 통해 사용자가 대화형 AI 튜터와의 학습과 대화에 있어 원활하게 진행되었는지, 대화형 AI 튜터와의 소통을 통해 학습 및 실력 향상에 긍정적인 영향을 미치는지 나타낸다. 강한 의인화가 사용성 관련 지표에서 더 높게 나타났다는 것은, 사용자와의 대화가 다양한 비언어적 의인화 요소 등으로 인해 교류가 된 인터페이스가 사용자에게 있어 학습이라는 의도와 자연스러운 사용자 입력과 발화를 유도함을 알 수 있다. 선행연구에서 대화형 AI 튜터와의 학습에 있어 장점으로 실시간으로 질문하면 바로 답을 얻을 수 있다고 인식하는 사용자의 특성에 따라, 강한 의인화에서 제공하는 빠른 피드백과 인터랙션이 사용성 지표에 긍정적인 영향을 미친 것으로 보인다(Chen et al., 2020).

Perceived Intelligence, Perceived Trust, Perceived Interactivity는 대화형 AI 튜터의 공감 및 감정적 상호작용 능력과 관련되어 있는 지표다. 이러한 지표들은 사용자의 참여, 집중, 시스템에 대한 태도에 영향을 미칠 수 있으므로, AI 기반 시스템의 개발에서 점점 더 중요해지고 있다(Feine et al., 2020). 이러한 지표들에서 강한 의인화가 약한 의인화, 텍스트 기반의 의인화보다 더 높은 점수를 나타낸 이유는 선행연구에서 인간과의 유사성만 단독으로 사용하면 의인화로 인식하고 상호작용하는 것에 부정적인 영향을 미치지만, 언어적 표현이나 인간과의 유사성에 비언어적 표현을 더한다면 이를 긍정적으로 인식할 수 있다는 결과와 일치한다(Seeger et al., 2021). 즉 회화 학습에서 대화형 AI 튜터와 대화하는 것에 있어서도 오직 의인화된 이미지만 제시하는 것보다는 다양한 비언어적, 언어적 표현들을 통해서 사용자와 교류하는 것이 대화형 AI 튜터를 인식함에 있어서 긍정적인 영향을 미침을 알 수 있었다.

강한 의인화가 Intention to Use에서 또한 높은 점수를 얻은 결과는, 앞선 FGI 연구에서 사용자들이 회화 학습 중 사용자가 대화형 AI 튜터의 사회적 존재감이 느껴지는 것을 원했으며, 대화형 AI 튜터와의 학습으로 인해 성취감과 같은 부가적인 요소를 선호한 것과 일치한 결과로 볼 수 있다. 즉, 텍스트 기반과 약한 의인화와 달리, 사용자의 발화에 대해서 즉각적인 피드백을 제시하고, 또 밈과 이모티콘을 이용하여 강한 사회적 존재감을 강조한 것이 Intention to Use 점수에 긍정적인 영향을 미쳤음을 알 수 있으며, 이는 이러한 비언어적, 언어적 요소가 추후 회화 학습에서 대화형 AI와 같은 새로운 기술을 도입하는 것에 있어 긍정적인 영향을 미칠 수 있을 것이라 기대된다.

5. Conclusion

본 연구에서는 대화형 AI를 활용한 외국어 회화 학습에서, 대화형 AI 튜터의 의인화를 통해 회화 학습 및 사용자 경험에 적합한 인터페이스를 제안하고자 했다. 본 연구에서는 최적의 인터페이스를 도출하기 위해, 사용자가 선호하는 대화형 AI의 시각적 의인화 정도와 속성을 도출하고자 하였으며 이와 관련된 인식을 조사하기 위해 대화형 AI를 사용한 경험이 있는 5명의 대학생을 대상으로 FGI 연구를 진행했다. 인터뷰를 통해 사용자는 실제 사람 이미지보다는 강한 캐릭터성을 가진 이미지를 선호하였으며, 특정 분야에 전문성을 가진 속성(의사, 선생님 등)보다 자신 또래의 이미지를 선호하는 것을 알 수 있었다. 또한 대화형 AI 튜터의 목소리와 의인화된 대화형 AI 튜터의 이미지와의 조화를 중시하였으며, 단순히 회화 학습뿐 아니라, 리워드나 스토리텔링과 같은 부가적인 요소들을 선호하였다. 본 연구에서 FGI 연구 대상으로 영어 회화 학습에 큰 필요성을 느끼고 AI 기술에 대한 수용성을 가지고 있는 20대 초반 대학생을 주요 검증으로 선정하였으나, 향후 연구에서는 연령과 사회경제적 배경이 다양한 참가자들을 포함시켜 사용자가 선호하는 의인화의 특성들을 추가로 검증할 필요가 있다.

FGI 연구를 통해 도출한 의인화의 특성을 통해 프로토타입을 제작하였다. 텍스트 기반의 비의인화 단계, 인간과 닮은 '외형'이 추가된

약한 의인화 단계, 이모티콘과 유머러스한 이미지 등 언어적, 비언어적 표현를 추가한 강한 의인화 단계라는 3가지 단계로 구분하여 프로토타입을 제작했다. 본 연구에서는 제작한 프로토타입이 실제로 외국어 회화 학습에 있어 유효한 결과를 나타내는지 확인하기 위해, 20명을 대상으로 실험을 진행하였다. 실험 지표로는 대화형 AI와 챗봇을 평가할 때 자주 사용되는 7가지 지표인 Perceived Ease of Use, Perceived Usefulness, Perceived Interactivity, Perceived Intelligence, Perceived Trust, Intention to Use를 사용하였다. 본 연구에서는 해당 지표들에 있어서 비의인화, 약한 의인화보다 강한 의인화일 때 더 높은 지표를 보일 것으로 예측하였다. 이를 통계적으로 검증하기 위해 ANOVA와 Tukey's test를 시행하였다.

실험 결과, Perceived Ease of Use, Perceived Usefulness, Perceived Interactivity, Perceived Intelligence, Perceived Trust, Intention to Use 6가지 지표에서 모두 텍스트 기반의 채팅, 약한 의인화 방식보다 강한 의인화에서 가장 높은 점수를 나타냈으며, 그 차이는 모든 지표에서 통계적으로 유의하게 나타났다. 또한 사후검정 결과 6가지 지표에서 약한 의인화, 텍스트 기반 방식은 통계적으로 유의한 차이가 없었던 것에 반해, 강한 의인화 방식은 2가지 방식 모두와 통계적으로 유의한 차이가 발생했다. 이를 통해 모든 가설이 지지됨을 확인할 수 있었다.

본 연구에서 제시하는 사용자 인터페이스는 향후 외국어 회화 학습 시에 활용될 수 있는 대화형 AI의 사용자 인터페이스를 설계할 때 활용될 수 있을 것이다. 대화형 AI는 시간이 부족한 학습자들을 위한 대안적 학습 방법으로 기대되며, 이에 따라 사용자의 대화에 대한 몰입을 돕는 사용자 인터페이스에 대한 연구가 필요하다. 본 연구에서는 의인화 요소에 있어 대화형 AI를 경험해본 사용자들이 선호하는 외형 및 속성을 비롯하여 대화 상대와 상호작용에 도움이 되는 언어적 표현, 이모지, 밈과 같은 트렌드에 반영된 비언어적 표현들을 결합하여 강한 의인화를 구현하고자 하였다. 따라서 이러한 강한 의인화에 해당하는 의인화적 표현들이 외국어 회화 학습에서의 대화형 AI에 대한 인터페이스를 구축할 때 사용자의 몰입 및 외국어 학습의 효율을 증대할 수 있는 요소로 기대된다.

향후 연구에서는 시험 점수와 같은 객관적인 점수를 통해 대화형 AI의 인터페이스에 대한 결과를 검증할 수 있을 것이다. 또한 FGI 연구에서 도출된 리워드, 스토리텔링과 같은 학습에 도움이 될 수 있는 부가적인 요소들에 대해서 실제로 학습의 몰입에 도움이 될 수 있는지 확인할 수 있다. 또한, 이번 실험에 참여자들보다 낮은 연령대나 대화형 AI에 익숙하지 않은 사용자들을 대상으로 실험을 진행할 수 있다.

Acknowledgement

This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1F1A1066216).

References

Adam, M., Wessel, M. and Benlian, A., AI-based chatbots in customer service and their effects on user compliance, *Electronic Markets*, 31(2), 427-445, 2021.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R. and Horvitz, E., Guidelines for human-AI interaction, *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems*, (pp. 1-13), 2019.

Ayedoun, E., Hayashi, Y. and Seta, K., 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, 2019.

Beattie, A., Edwards, A.P. and Edwards, C., A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication, *Communicating Artificial Intelligence (AI)*, Routledge, 41-59, 2020.

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

Bibauw, S., François, T. and Desmet, P., Dialogue systems for language learning: Chatbots and beyond, *The Routledge Handbook of Second Language Acquisition and Technology*, Routledge, 121-135, 2022.

Brachten, F., Kissmer, T. and Stieglitz, S., The acceptance of chatbots in an enterprise context–A survey study, *International Journal of Information Management*, 60, 102375, 2021.

Chandra, S., Shirish, A. and Srivastava, S.C., To be or not to be… human? Theorizing the role of human-like competencies in conversational artificial intelligence agents, *Journal of Management Information Systems*, 39(4), 969-1005, 2022.

Chen, H.L., Vicki Widarso, G. and Sutrisno, H., A chatbot for learning Chinese: Learning achievement and technology acceptance, *Journal of Educational Computing Research*, 58(6), 1161-1189, 2020.

Cheng, X., Zhang, X., Cohen, J. and Mou, J., Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms, *Information Processing & Management*, 59(3), 102940, 2022.

Choi, J.G. and Kim, M., The usage and evaluation of anthropomorphic form in robot design, 2008.

Davis, F.D., Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*, 319-340, 1989.

De Graaf, M.M. and Malle, B.F., How people explain action (and autonomous intelligent systems should too), *2017 AAAI Fall Symposium Series*, 2017.

Dillon, T., Creating an Alexa-enabled textbook exercise: An easy approach to custom automatic speech recognition application, *KOTESOL Proceedings*, 259, 2018.

Etemad-Sajadi, R., The impact of online real-time interactivity on patronage intention: The use of avatars, *Computers in Human Behavior*, 61, 227-232, 2016.

Feine, J., Morana, S. and Maedche, A., Designing Interactive Chatbot Development Systems, *ICIS*. 2020.

Fryer, L.K., Ainley, M., Thompson, A., Gibson, A. and Sherlock, Z., Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners, *Computers in Human Behavior*, 75, 461-468, 2017.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A. and Sasse, M.A., The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 529-536), 2003.

Gayashan, S.P. and Samarasinghe, S., Factors Affecting the Intention to Use AI-Based Chatbots in Learning, *2024 International Research Conference on Smart Computing and Systems Engineering (SCSE)* (Vol. 7, pp. 1-9), IEEE, 2024.

Ha, L. and James, E.L., Interactivity reexamined: A baseline analysis of early business web sites, *Journal of Broadcasting & Electronic Media*, 42(4), 457-474, 1998.

Haugeland, I.K.F., Følstad, A., Taylor, C. and Bjørkli, C.A., Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design, *International Journal of Human-Computer Studies*, 161, 102788, 2022.

Ho, A., Hancock, J. and Miner, A.S., Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot, *Journal of Communication*, 68(4), 712-733, 2018.

Interaction Design Foundation – IxDF, What are User Flow, https://www.interaction-design.org/literature/topics/user-flows (retrieved June 24, 2024).

Jeon, J., Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives, *Computer Assisted Language Learning*, 37(1-2), 1-26, 2024.

Krueger, R.A., Focus Groups: A Practical Guide for Applied Research, Newbury Park. CA, Sage, 1988.

Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N. and Gadgil, K., Conversational AI: An overview of methodologies, applications & future scope, *2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, (pp. 1-7), 2019.

Li, M. and Suh, A., Anthropomorphism in AI-enabled technology: A literature review, *Electronic Markets*, 32(4), 2245-2275, 2022.

Lee, S., Lee, N. and Sah, Y.J., Perceiving a mind in a chatbot: effect of mind perception and social cues on co-presence, closeness, and intention to use, *International Journal of Human-Computer Interaction*, 36(10), 930-940, 2020.

Moussawi, S. and Koufaris, M., Perceived intelligence and perceived anthropomorphism of personal intelligent agents: Scale development and validation, 2019.

Moussawi, S., Koufaris, M. and Benbunan-Fich, R., How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents, *Electronic Markets*, 31(2), 343-364, 2021.

Mohd Tuah, N., Wills, G. and Ranchhod, A., The characteristics and application of anthropomorphic interface: A design spectrum, 398-402, 2016.

Munnukka, J., Talvitie-Lamberg, K. and Maity, D., Anthropomorphism and social presence in Human-Virtual service assistant interactions: The role of dialog length and attitudes, *Computers in Human Behavior*, 135, 107343, 2022.

Murano, P., Gee, A. and Holt, P.O., Evaluation of an anthropomorphic user interface in a travel reservation context and affordances, *Journal of Computing*, 3(8), 2011.

Neri, A., Cucchiarini, C. and Strik, H., Feedback in computer assisted pronunciation training: technology push or demand pull?, 2002.

Pillai, R. and Sivathanu, B., Adoption of AI-based chatbots for hospitality and tourism, *International Journal of Contemporary*

Hospitality Management, 32(10), 3199-3226, 2020.

Pillai, R., Sivathanu, B., Metri, B. and Kaushik, N., Students' adoption of AI-based teacher-bots (T-bots) for learning in higher education, *Information Technology & People* ahead-of-print, 2023.

Pillai, R., Ghanghorkar, Y., Sivathanu, B., Algharabat, R. and Rana, N.P., Adoption of artificial intelligence (AI) based employee experience (EEX) chatbots, *Information Technology & People*, 37(1), 449-478, 2024.

Qiu, H., Li, M., Shu, B. and Bai, B., Enhancing hospitality experience with service robots: The mediating role of rapport building, *Journal of Hospitality Marketing & Management*, 29(3), 247-268, 2020.

Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G.M., Brunskill, E. and Landay, J.A., Englishbot: An ai-powered conversational system for second language learning, *26th International Conference on Intelligent User Interfaces* (pp. 434-444), 2021.

Seeger, A.M., Pfeiffer, J. and Heinzl, A., Texting with humanlike conversational agents: Designing for anthropomorphism, *Journal of the Association for Information Systems*, 22(4), 8, 2021.

Strohmann, T., Siemon, D., Khosrawi-Rad, B. and Robra-Bissantz, S., Toward a design theory for virtual companionship, *Human-Computer Interaction*, 38(3-4), 194-234, 2023.

Tarhini, A., Masa'deh, R.E., Al-Busaidi, K.A., Mohammed, A.B. and Maqableh, M., Factors influencing students' adoption of e-learning: a structural equation modeling approach, *Journal of International Education in Business*, 10(2), 164-182, 2017.

Van Pinxteren, M.M., Pluymaekers, M. and Lemmink, J.G., Human-like communication in conversational agents: a literature review and research agenda, *Journal of Service Management*, 31(2), 203-225, 2020.

Wang, X., Pang, H., Wallace, M.P., Wang, Q. and Chen, W., Learners' perceived AI presences in AI-supported language learning: a study of AI as a humanized agent from community of inquiry, *Computer Assisted Language Learning*, 1-27, 2022.

Yang, H., Kim, H., Lee, J.H. and Shin, D., Implementation of an AI chatbot as an English conversation partner in EFL speaking classes, *ReCALL*, 34(3), 327-343, 2022.

Yu, C.E., Humanlike robots as employees in the hotel industry: Thematic content analysis of online reviews, *Journal of Hospitality Marketing & Management*, 29(1), 22-38, 2020.

Zhou, P., Xie, Y. and Liang, C., How to increase consumers' continued use intention of artificial intelligence voice assistants? The role of anthropomorphic features, *Electronic Markets*, 33(1), 60, 2023.

Author listings

Ga Young Yoo: gayoungyoo.y@gmail.com
Highest degree: Undergrad. Student, Department of Art Studies, Hongik University
Position title: Undergrad. Student, Department of Art Studies, Hongik University

Areas of interest: Human-computer Interaction, Usability Engineering, AI User Interface

Seo Yeon Kim: yeonmm04@g.hongik.ac.kr
Highest degree: Undergrad. Student, Department of Visual communication design, Hongik University
Position title: Undergrad. Student, Visual communication design, Hongik University
Areas of interest: Human-computer Interaction, UX design and management, AI User Interface

Kyungdoh Kim: kyungdoh.kim@hongik.ac.kr
Highest degree: PhD., School of Industrial Engineering, Purdue University
Position title: Associate Professor, Department of Industrial & Data Engineering, Hongik University
Areas of interest: Human-computer Interaction, Usability Engineering, UX design and management