Master of Science Thesis

Examining the Effects of Cognitive Engagement and Chatbot Humanness on Learning

인지적 참여도와 챗봇의 휴먼니스가
학습에 미치는 영향에 대한 탐구

August 2023

Graduate School of Convergence Science and Technology
Seoul National University
Intelligence & Information Major

Jiwon Kim

Examining the Effects of Cognitive Engagement and Chatbot Humanness on Learning

Advisor Gahgene Gweon

Submitting a Master of Science Thesis

August 2023

Graduate School of Convergence Science and Technology
Seoul National University
Intelligence and Information Major

Jiwon Kim

Confirming the M.Sc. Thesis written by
Jiwon Kim
August 2023

Chair	Sowon Hahn	(Seal)
Vice Chair	Gahgene Gweon	(Seal)
Examiner	Joonhwan Lee	(Seal)

Abstract

The use of text-based disembodied chatbots in education has recently gained attention in educational fields. To develop and design chatbots for educational purposes, two types of interactions—learner and educator, user and product— have to be simultaneously considered. As such, this study aims to examine the impact of a learning experience-related factor (i.e., cognitive engagement in learning) and a user experience-related factor (i.e., cognitive determinants of chatbot humanness) on learning outcomes and motivation.

Four versions of chatbots (i.e., constructive & humanized, constructive & non-humanized, active & humanized, active & non-humanized chatbots) were designed, based on the ICAP framework and the CMD framework. Constructive chatbots lead learners to (1) explain or summarize concepts in the video/simulation in one's own word; (2) justify my opinion based on what s/he learned; (3) infer new information from what was explicitly taught; (4) generate predictions on a new case that was not addressed during learning. In contrast, active chatbots require learners to fill in the blank of the full sentences or select a correct answer among the various options, without the need to generate a new output or infer new information. Humanized chatbots were designed by using six cognitive determinants of perceived humanness: (1) provide a visual representation of humanlike character, (2) identify a chatbot's role and name as humanlike, (3) use social dialogues, (4) use informal and casual verbal style, (5) use emojis, and (6) use stickers that have facial expressions.

Our first and second hypotheses were set to examine the main effect of the cognitive engagement mode of learning and cognitive determinants of chatbot humanness, respectively. The third hypothesis was set to investigate the interaction effect of the two independent variables. A mixed method approach was taken to examine the three hypotheses by comprehensively analyzing quantitative and qualitative data. We measured learning outcomes by pre/post-tests and learning motivation by post-survey and interview. Using such measurements, we conducted two-way ANOVA and Rank-sum Transformation ANOVA. The contributions of our study are four-fold: (1) We designed our educational chatbot based on the concepts of learning experience design to encompass both the instructional design element and the user experience element; (2) Despite the learners' lacking background knowledge, constructive chatbot were more effective in improving learning outcomes than active chatbots; (3) Different from our prediction, which was based on previous literature, non-humanized chatbots were more effective in learning outcomes than humanized chatbots; (4) When using constructive chatbots, cognitive determinants of chatbot humanness gave different impacts on two types of learning motivation: positive impact on tension-pressure and negative impact on perceived competence.

Keyword: #Experimental research #Educational chatbot #ICAP framework #Cognitive engagement in learning #Humanness of chatbot #Mixed method

Student Number: 2021-23814

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

A chatbot, or a conversational agent, is an agent that mimics human-to-human communication by giving an instant answer to users via either a text-based or an audio-based conversation (Jain et al., 2018; Rubin et al., 2010). Two types of conversational agents exist: embodied conversational agents and disembodied conversational agents. Embodied conversational agents are virtual, three-dimensional characters displayed on device screens and interact with people through natural speech. In contrast, disembodied conversational agents are not visually represented and thus depend solely on text-based or audio-based conversation, facilitating easy and fast development and increasing accessibility.

Along with the prevalence of disembodied conversational chatbots in daily life (Araujo, 2018), the use of text-based disembodied chatbots in education has recently gained attention in educational fields as a way to enhance anywhere and anytime education and to support personalized learning (Smutny & Schreiberova, 2020). First, learners can use educational chatbots (EC) without being limited by location and time (L. Fryer & Carpenter, 2006; Haristiani, 2019; Jia, 2009). Second, when using Ecs, learners can learn in user interfaces with which they are already familiar, such as Telegram, Facebook Messenger, and Slack (Jain, Kumar, Kota, et al., 2018; Skjuve & Brandtzæg, 2018). Such familiarity can lessen learners' effort that might have been required to get used to learning with Ecs. Third, by having back-and-forth dialogues with Ecs, learners can feel like they are talking to a human instructor, which can enhance chatbot users' learning experience (N.-Y. Kim et al., 2019; Yin et al., 2021). Fourth, Ecs can provide personalized learning experiences that compensate for the limitations of traditional educational environments, which may not fully meet every student's needs. As such, Ecs have the potential to facilitate more effective and efficient learning experiences for learners (Deveci Topal et al., 2021).

To develop and design chatbots for educational purposes, two types of interactions—learner and educator, user and product— have to be simultaneously considered. The interaction between a learner and EC as an educator relates to creating the learning experience that enables a learner to achieve a desired learning outcome (Floor, 2023). The interaction between a user and EC as a product pertains to shaping the user experience of using a product. Learning experience design, or LXD, is the process of creating learning experiences that enable the learner to achieve the desired learning outcome in a human-centered and goal-oriented way. Since LXD encompasses both aspects of traditional instructional design and user experience design, it can be a good foundation for investigating how to design educational chatbots to maximize their learning effects. As such, this study aims to examine the impact of a learning experience-related factor (i.e., cognitive engagement in learning) and a user experience-related factor (i.e., perceived humanness) on chatbot-based learning.

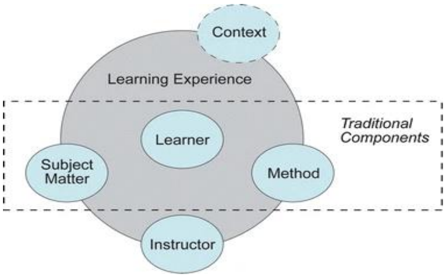

Figure 1. Learning experience, which is the transaction between a learner and the components of instructional environments

As shown in Figure 1, the learning experience describes the transactions that take place between individual learners and the instructional environment. Specifically, the learning experience includes how the learner feels about, engages with, responds to, influences, and draws from the subject matters, instructional methods, instructors, and learning contexts (Parrish, 2009). Concerning the subject matters, Ecs have been utilized for supporting language learning (Ji et al., 2023; Pham et al., 2018; Wang et al., 2023), computer science education (Benotti et al., 2014), science education (Deveci Topal et al., 2021), and mathematics education (Lee & Yeo, 2022; Ruan et al., 2020). Regarding the instructional methods and instructors, Ecs have taken various roles, such as companions, assistants and mentors, to improve learner-instructor interaction (S. Kim et al., 2020; Lin & Chang, 2020; H. Nguyen, 2022). Most Ecs are set as learning partners who are more experienced and competent than learners, guiding learners to do specific learning tasks, such as problem-solving (Billett, 2012; Hwang & Chang, 2021).

However, previous works have not concentrated on considering learner engagement in learning subject matters. Learner engagement has three aspects: cognitive, emotional, and behavioral. Among them, cognitive engagement is known to have a close relationship with learning effects, such as improvement in learning outcomes. According to the ICAP framework (Chi & Wylie, 2014), as learners are more cognitively engaged in learning, the learners are expected to show better learning outcomes than those who are engaged less in learning. On the other hand, such expectations grounded on the ICAP framework might only be consistent when learners have the background knowledge to the extent that even low cognitive engagement in learning can be a significant burden (Chi and Wylie, 2014). Thus, it is necessary to examine which cognitive engagement mode of learning can positively impact chatbot-based learning where learners need more previous knowledge on subject matters.

In addition to considering how learners are cognitively engaged in learning, it is also essential to consider how learners perceive instructors. Especially when the instructors are not human but technological systems such as chatbots, educational chatbots can be regarded as a product and the learner as an end-user. The users' perception of chatbots, such as perceived humanness, has been broadly investigated in Human-Computer Interaction fields due to its impact on user experiences (Go & Sundar, 2019; Y. Jiang et al., 2023; Rhim et al., 2022). The *humanness of chatbot* is defined as chatbots having humanlike traits in terms of behavior, personality, conversational styles, or appearances (Diederich et al., 2022; Rapp et al., 2021). According to Seeger et al. (2021)'s framework of determinants of chatbot humanness perceived by users, cognitive determinants can influence various aspects of user experience, including trust in chatbots (Chattaraman et al., 2019), user satisfaction (H. Jiang et al., 2022), and perceived ease of use (Sheehan et al., 2020). Meanwhile, much research has been conducted to examine whether using cognitive determinants of chatbot humanness positively impacts learning with embodied conversational agents (Dai et al., 2022; Li et al., 2022; Woo, 2009). However, such results were not extensively examined in the context of 'text-based,' 'disembodied' educational chatbots. Thus, examining whether using cognitive determinants of chatbot humanness in text-based educational chatbots can lead to better learning is necessary. Meanwhile, considering that each component of the learning experience interacts with each other holistically, *cognitive* engagement in learning will be influenced by learners' perception of Ecs, which *cognitive* determinants of chatbot humanness can manipulate. Thus, we set our research question as follows: *RQ*−*Do cognitive* *engagement in learning and perceived humanness of chatbot have significant main/ interaction effects on learning outcomes and motivation of chatbot users?*

The contribution of this research is four-fold. First, we showed the effectiveness of grounding on LXD (or learning experience design), that can encompass both elements of traditional instructional design and user experience design. Second, our study results validated that the hierarchy relationship between the constructive and the active mode of learning within the ICAP framework is valid in educational chatbot contexts. Thirdly, we showed that cognitive determinants of perceived humanness can be utilized to humanize educational chatbots and which determinants are recommended to use In future research. Fourthly, we showed that the non-humanized design of text-based chatbots can bring better learning outcomes than humanized chatbots. Lastly, our study is the first to prove the interaction effect between humanness of chatbots and learning engagement on learning motivation.

The rest of this paper Is organized as follows. In chapter two, we will review previous works about educational chatbots, cognitive engagement in learning, and the humanness of chatbot. Chapter three describes how the four versions of our study's chatbots are designed. Concerning the experiment, chapters four, five, and six describe the method, results, and discussion, respectively. In chapter seven, the conclusion, this study's limitations, and suggestions for future study are described.

Chapter 2. Literature Review

2.1. Chatbots used for Educational Purposes

Chatbots have been utilized to support learning various subject matters. For language education, many types of chatbots help learners understand grammar rules (Jia et al., 2012; N.-Y. Kim et al., 2019), practice conversation (Ayedoun et al., 2015; L. K. Fryer et al., 2017), and memorize vocabulary (Chen et al., 2020; Jia et al., 2012). Chatbots have been used for STEM education, such as mathematical problem-solving (H. D. Nguyen et al., 2019), figuring out knowledge related to scientific concepts (Deveci Topal et al., 2021) and practicing programming (Farah et al., 2022).

Chatbots have been used both for supporting online learning and in-person classroom learning. Fidan & Gencel (2022) showed the positive impact of using their chatbot that offered feedback on students' learning. Students who interacted with this chatbot while taking pre-recorded lectures showed a higher learning motivation and achievement level than those who did not. Winkler et al. (2020) showed the positive impact of using their chatbot that had learners solve questions related to the lectures and checked learners' understanding of the lectures. Students who interacted with a chatbot while watching video instructions were better at recalling the contents of the lectures and applying such contents to other situations than those who did not use the chatbot. Deveci Topal et al. (2021) developed a chatbot that answered learners' queries and recommended additional learning resources. A group of students who used this chatbot while taking a synchronous online science class had a more positive learning experience than the other group who did not use the chatbot. Jeon (2022) designed a chatbot that took a role of a native English-speaking conversation partner for those who were not native English speakers. Students who used this chatbot showed a higher level of learning motivation than those who did not. Lin & Chang (2020) developed a chatbot to support improving learners' writing skills. A group of learners used this chatbot within the curriculum of the writing course, and their disciplinary writing skills showed a significant improvement.

2.2. Cognitive Engagement in Learning and the ICAP Framework

2.2.1. ICAP Framework and Chatbots for Educational Purposes

The ICAP framework categorizes four cognitive engagement modes of learning based on the observable overt behaviors that learners show during the learning process. The modes are referred to as 'interactive,' 'constructive,' 'active,' and 'passive' (Chi & Wylie, 2014). In the passive mode, learners merely receive and store knowledge and are expected to be able to recall such knowledge. In the active mode, learners manipulate and integrate pieces of knowledge they learned. The learners are expected to be able to apply such knowledge in a new context. In the constructive mode, learners generate new knowledge by inferring from what they have already learned. The learners are expected to transfer knowledge, which is more cognitively engaging than mere recall and manipulation of knowledge. In the interactive mode, more than two learners interact with each other and co-create new knowledge. To confirm that the learners show the interactive mode of cognitive engagement in learning, it is necessary that all the learners should simultaneously involve themselves in constructive mode.

Several works developed chatbots for various educational purposes and tried to show the benefits of such chatbots based on the ICAP framework. Most of such works assumed that the interaction between a chatbot and a learner leads to the interactive mode of cognitive engagement in learning (Wambsganss et al., 2021; Winkler et al., 2020; Xu et al., 2023). For example, Winkler et al. (2020) and Winkler et al. (2021) developed a chatbot that assists learners with participating in video lectures. It was demonstrated that the positive effect of using such a chatbot was enhanced when the chatbot used dynamic scaffolding, an instructional method that activated an interactive mode of learning. Wambsganss et al. (2020, 2021) designed an educational chatbot that supports learners in improving their argumentation skills. They analyzed that the deep interaction between such a chatbot and a learner induced an interactive cognitive engagement mode, leading to higher learning outcomes. Similarly, Xu et al. (2023) designed a voice-based chatbot that teaches mathematical language. They claimed that while a learner interacted with such chatbots, they were allowed to expand their knowledge. Such expansion of knowledge indicated an interactive mode of cognitive engagement in learning, leading to higher learning outcomes.

2.2.2. Constructive Mode of Learning

According to the ICAP framework (Chi & Wylie, 2014), exhibiting constructive learning behavior means that the learner creates additional outputs beyond what is learned from the learning material. This output should contain new information beyond the information given in the learning activity. Constructive learning promotes knowledge transfer because the structure of knowledge is substantially changed. Specifically, the acquired knowledge can be applied to new contexts or situations that have not been encountered, facilitating the acquisition of new knowledge concepts.

Constructive activities involve inferring the knowledge of various types of learners. ‘Inference’ includes inducing, deducing, and abducing. In the process of making such reasoning, various aspects of learning can appear as follows: correcting the concept you have misunderstood or modifying the structure of the existing knowledge; extracting and linking parts of articles from various sources; combining the contents of different paragraphs within the same article and summarizing them in a single sentence; comparing and contrasting articles in different parts with each other; creating new texts or new analogies based on sentences in learning materials.

Suppose a learner is asked to self-explain her solution to a set of problems. when a learner makes an unspecified inference in a given problem and writes it down as a sentence, it can be classified as constructive. For another instance of solving worked-out examples, if a learner solves a physics problem by drawing a picture not included in the given step-by-step solution, it can be classified as constructive because a new inference was made. Examples of learning activities that exhibit a constructive mode of engagement can be externalized as follows: when taking a lecture, speaking out loud while reflecting on one’s learning, drawing a concept map, and asking questions voluntarily; when reading articles, self-explaining, integrating between multiple texts, and taking notes in your own language; when watching educational video clips, explaining the concept in the video and comparing/contrasting with prior knowledge or other learning skills.

2.2.3. Active Mode of Learning

Learners who display active activities emphasize specific parts of a given learning material. This behavior may be an attempt by the learner to activate his or her prior knowledge and use it to understand new knowledge. When a related schema is activated, newly learned information can be assimilated into that schema. Moreover, learners can fill that gap. Active learning activates prior knowledge to help you become more proficient in learning knowledge in a new context. Therefore, it can be said that active engagement results In a significant level of knowledge completion.

Exhibiting active learning behaviors means learners manipulate what they learn from learning materials. Such manipulation may include actions such as pointing out the problem the learner is reading or solving, revisiting the learning material, highlighting parts that are considered important, or selecting the correct answer in a multiple-choice question and justifying it. Examples of learning activities that exhibit an active mode of engagement are as follows: when taking a lecture, repeating or rehearing learning materials, following the step-by-step steps of a model answer, or arranging words included in learning materials on a learner's own note; when reading an article, underlining important parts or marking them in a different color, writing down the text without making any change or just shortening its content; when watching an educational video, controlling the learning process by taking actions such as stopping, playing, fast-playing, and pressing rewind.

Unlike passive learning, active learning is different because it performs actions that show high concentration when manipulating knowledge. For example, when a learner points to parts of a given article but mindlessly and mechanically, such behavior will be classified as not active but passive. For another example, underlining important parts when reading a textbook is a process of overtly concentrating and acquiring knowledge. However, habitually and repeatedly underlining all sentences without thinking will be considered passive learning. In other words, a precondition for active learning is that a learner is externally focusing on learning and manipulating learning materials.

2.3. Cognitive Determinants of Chatbot Humanness

Seeger et al. (2021) suggested a framework of anthropomorphic (in other words, humanlike) conversational agent design. They clarified that three dimensions impact the perceived humanness of a chatbot: cognitive, motivational, and dispositional. Cognitive determinants pertain to how humanlike computer interfaces are designed or perceived. Cognitive determinants are manipulated through changing user interfaces, which are technological parts of text-based chatbots. Motivational determinants affect users' motivation to perceive non-human objects depending on the task. For example, merely handling banking tasks is close to a computer-like task, so users are not motivated to perceive the chatbot as human. However, situations like discussion are closer to humanlike tasks, so users try to perceive the chatbot more like a human. Depositional determinants relate to individual differences, such as cognitive abilities, personality traits, demographics, and cultural background.

In the experimental setting of the present study, both the motivational and the depositional determinants were controlled. This is because the task provided by our chatbots is consistently applied across all conditions as problem-solving, and the context is set to AI foundational knowledge education for non-experts. As such, the following subsections will explore the existing literature pertaining to three critical design dimensions of cognitive determinants of chatbot humanness: human-like identity cues, verbal expression cues, and non-verbal expression cues. Based on (Epley et al., 2007)'s work, human identity cues given to non-human identity are stable features that do not change drastically and are set up before the user-computer interaction, impacting users' perception of non-human entities, such as chatbots. Verbal and nonverbal expression cues are behavioral features that change during user-computer conversation.

2.3.1. Human identity cues

According to Seeger et al. (2021), human identity cues of a chatbot can be expressed through the following two settings: (1) by being visually represented in humanlike figures, such as a real human and a humanlike character; (2) by being given properties that are unique to humans, such as age, gender, name, ethnicity, job, and personality.

Humanlike visual representation – Various outcomes of perceiving chatbots as similar to humans have been explored, including service satisfaction (Gnewuch et al., 2018), brand perception (Araujo, 2018), and learning gain (Jin et al., 2010). The presence of humanlike visual cues in chatbot design has significant positive effects on trust, purchase intention, word of mouth, and satisfaction with the shopping experience, as demonstrated by Konya-Baumbach et al. (2023). According to Sundar (2008), humanlike visual cues are likely to trigger a sense of humanness in users, leading them to perceive chatbots as human and engage with them socially. This notion is supported by studies conducted by Y. Kim & Sundar (2012) and Gong & Nass (2007). Additionally, highly anthropomorphic visual cues, such as a human figure, increase the salience of the other person in the interaction, as argued by Go & Sundar (2019), implying the existence of another individual. Moreover, Rocca & McCroskey (1999) suggest that anthropomorphic visual cues of an online chat agent enhance the perception of similarity (homophily) between the agent and the user compared to non-humanlike figures. However, it is important to note that humanized chatbot design can also induce feelings of uncanniness in users, as discussed by Gnewuch et al. (2018) and Wuenderlich & Paluch (2017).

Demographic information – Identifying the demographic information of chatbots as that of humans can make users perceive chatbots as humanlike. (Araujo, 2018) used a human name, Emma, to make a disembodied chatbot that responds to questions from online shopping customers more humanlike, while a non-humanlike chatbot was given the name ChatBotX, a name that is not commonly given to humans. Qiu & Benbasat (2010) explored the impact on user experience when the race and gender of an embodied conversational agent, who assists in recommending purchasable goods, match those of the user. When the agent's race matched that of the user, the user perceived the agent as more sociable, enjoyable, and useful to interact with, compared to when there was no match. The congruence of gender between the agent and user did not significantly influence such perceptions. Benbasat et al. (2010) investigated the influence of the gender and race match between an interactive agent recommending desired items and the user on the user's perception of social presence. It was observed that female users perceived a higher level of social presence when the agent was of the same gender compared to when it was not, while male users did not exhibit a significant impact. Harrington & Egede (2023) examined the impact of a chatbot's race and age, designed for the health information-seeking behavior of older Black adults, being identical to the user's, on user experience. The findings revealed that the chatbot's likeness with the user in terms of race and age had a significant influence on the feelings of trust and comfort with chatbots.

2.3.2. Verbal Expression Cues

Social dialogue includes greetings and non-task-related questions, such as small talk. The following works in HCI fields showed that social dialogues significantly impact various aspects of user experiences. T. Bickmore & Cassell (2001) examined the differences in user experiences depending on the chatbot's use of small talk. There was a significant effect on the users' trust in the chatbot due to the Interaction between the chatbot's use of small talk (social-oriented chatbot and task-oriented chatbot) and the user's disposition (extroverted and introverted). To elaborate, for extroverted individuals, the use of small talk by the chatbot had a significant influence on enhancing trust in the chatbot. In contrast, for introverted individuals, trust was improved with the task-only chatbot. T. W. Bickmore & Picard (2005) found that a chatbot utilizing social-emotional and relationship-building skills (i.e., a relational chatbot) led users to show more respect, have a greater liking, and display more trust towards the chatbot, compared to when a chatbot that did not use such skills (i.e., a task-oriented chatbot) was used. Chattaraman et al. (2019) explored the impact of conversation styles (social-oriented or task-oriented) employed by a conversational agent assisting with shopping on perceived trust in the ability of chatbots, perceived trust, perceived two-way interactivity, and perceived synchronous interactivity. In the social-oriented conversation style, the agent had an informal conversation with a user through small talk, questions, exclamatory feedback, and encouragement. It was revealed that older adults with high internet usage skills rated user-agent interactivity higher and had more trust in the website when communicating through a social-oriented conversation than a task-oriented conversation style. Hu et al. (2018) developed a chatbot that responds to online shopping customers on social media by analyzing the tone of the user's utterance to predict the emotional state underlying the tone and provide empathetic responses. Through such empathetic social dialogue, the chatbot made users feel as if they were interacting with a human-like agent, even more so than when interacting with a human agent.

Verbal style includes self-references, active voice, variability of syntax and words, and formality of conversation. Chattaraman et al. (2019) employed a humanlike verbal style, such as informal conversation, to allow a chatbot as an online shopping assistant to induce more human-like interactions. Specifically, informal conversation entails consistently using small talk, questions, exclamatory feedback, encouragement, and providing functional guides and information. For instance, they added variation to the verbal style by using exclamation marks at the end of sentences, using suggestive sentences instead of imperative ones (for example, "Can you do this for me?" instead of "Please do something.") and using emotional reactions like "Great!" In Sah & Peng (2015)'s work showed that using humanlike language by chatbots can increase social perception and facilitate chatbot users to leave a review on the website. Personal style was designed as a conversational form in complete sentences, with active voices and personal pronouns in all possible places in the questions (e.g., "How many times in the last year did you feel guilty after drinking?"). In contrast, impersonal style questions used passive voice and avoided referring to respondents. They were phrased using nominalized terms and abbreviations (e.g., "No. of times in the last year having a feeling of guilt after drinking").

2.3.3. Nonverbal Expression Cues

To humanize chatbots in terms of nonverbal expressions, we will use two types of cues: emojis and stickers. Within computer-mediated communication (CMC), emojis and stickers are typically categorized as nonverbal expressions. Such nonverbal cues can also be incorporated into text-based chatbots. Using emojis and stickers in computer-mediated communication (CMC) can complement the absence of nonverbal cues in text-based conversations (Beattie et al., 2020; Janssen et al., 2014; Tang & Hew, 2019). Emoji is a Unicode-matched text with an image format that visualizes emotions through facial expressions, hand gestures, icons, and objects. Sticker is a new type of emoji as the advance of CMC technology has enabled users to send various images easily (Zhou et al., 2017). While emoji is narrowly defined as cannot be sent separately from the text, a sticker is defined as a static or animated image that is sent independently from the text (Cha et al., 2018).

In-text emojis – Emojis are always available regardless of the type of messenger since they are matched with Unicode (UTF-8). Several instant messaging applications such as Facebook Messenger, Instagram Messenger, WhatsApp, and Naver Line provide or sell stickers each company designed. In the Facebook app and Instagram apps, as well as Apple's iPhone and Samsung's Android-based smartphones, users can create personalized stickers. They can customize their characters' eye colors, face shapes, items of clothing, accessories, and so on. Specifically, such stickers are called Animoji (the combination of animal and emoji) and Memoji (the combination of 'me' and emoji).

Emojis and stickers have shown various positive effects in CMC contexts. First, it can enhance users' perceived intimacy toward virtual agents (Roberta et al., 2020). Also, emojis and stickers that show positive emotions could induce users' positive feelings during CMC (Elder, 2018). In educational contexts, several works investigated how emojis and stickers can impact the learning experience (Bai et al., 2019). Regarding learners' emotions, when an instructor used emojis when communicating with their students via messengers, students felt a higher level of intimacy and friendliness toward the instructor compared to when the instructor used text only (Farah et al., 2021; Moffitt et al., 2020; Vareberg et al., 2022). Regarding learning efficacy, it was reported that emojis could help learners understand learning materials easily (Brody & Caldwell, 2019; Dunlap et al., 2016; Fane et al., 2018). Recent works tried to make learners feel a higher level of humanness toward an EC using emojis (Lee & Yeo, 2022; H. Nguyen, 2023). However, less effort was made by using stickers. Additionally, to the best of our knowledge, there has been no research to investigate the interaction between the cognitive engagement level of learning based on the ICAP framework and the humanness of chatbots.

Stickers with facial expressions – Nonverbal cues in direct human interaction significantly contribute to relationship development, serving various functions such as information conveyance, interaction regulation, emotional expression, self-introduction, greetings, and an indication of interpersonal attitudes (Argyle, 2013). Such nonverbal behaviors, including facial expressions, are also integral to Embodied Conversational Agents (ECAs), which use them alongside speech and text as communication channels (Radziwill & Benton, 2017). The facial expressions of these agents are known to affect users' perceptions and behaviors toward chatbots (Milcent et al., 2022). For instance, a study by Baylor & Kim (2009) assessed the impact of animated pedagogical agents' facial expressions on students' perceptions of the agent's persona, their attitudes towards learning content, and the overall learning process. The findings revealed a significant correlation between agent facial expressions, gestures, and enhanced perceptions and learning outcomes (Cowell & Stanney, 2005). Furthermore, nonverbal cues like facial expressions, eye contact, and paralanguage can foster a sense of trust between the participant and the virtual character. Studies show that pedagogical conversational agents employing nonverbal cues can significantly boost learning compared to agents communicating solely via text (Cook et al., 2017).

In disembodied text-based chatbots, it is limited for chatbots to use gestures because the chatbots do not have embodied avatars or characters through which they express gestures and facial expressions. However, by using stickers, which can be sent separately from textual sentences, gestures, and facial expressions can be expressed.

2.4. Research Gap and Hypothesis

In this section, the gap between previous studies and the present study will be described. We will set our research questions and corresponding hypotheses based on such research gaps. Figure 2 shows how the hypotheses of the present study are organized.

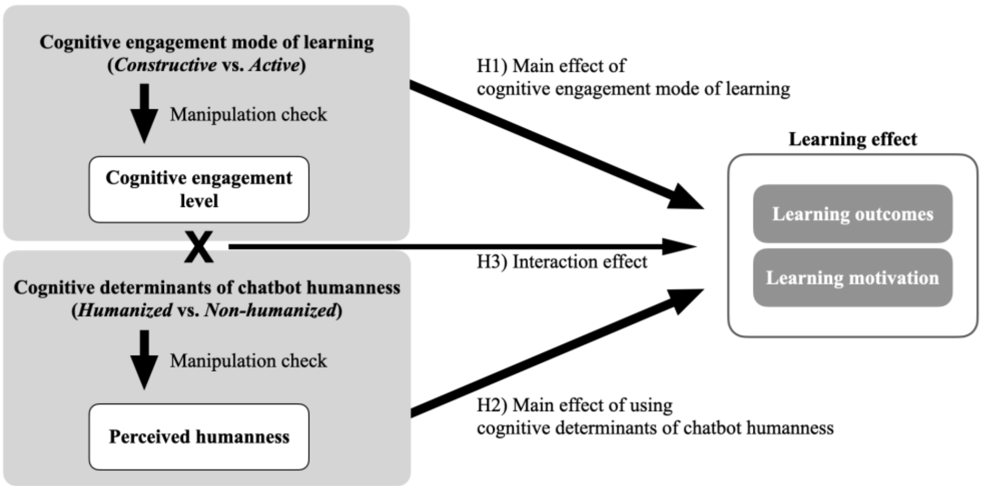

Figure 2. Three hypotheses to examine the main/ interaction effects of the two independent variables (i.e., cognitive engagement mode of learning and cognitive determinants of chatbot humanness) on the two dependent variables (i.e., learning outcomes and motivation)

2.4.1. H1: Main Effect of Cognitive Engagement Mode of Learning

In the previous works that were reviewed in *subsection 2.2.1. Educational Chatbots and the ICAP Framework*, it was insisted that the interaction between a learner and a chatbot led to the *interactive* mode of learning, which led to learning effects. However, according to the ICAP framework, an *interactive* mode of learning is achieved when both learners are engaged in the *constructive* mode of learning. Strictly speaking, it is hard to say that chatbots can be cognitively engaged in a constructive mode of learning unless such chatbots are based on generative AI models. Large Language Model (LLM)-based chatbots can infer new knowledge and information from given learning materials. However, due to the concerns of LLM providing inappropriate content to learners, most works reviewed in the previous section used rule-based or non-LLM chatbots. Such chatbots generate responses based on the programming logic already set by developers. Thus, such chatbots do not create new knowledge schema or make inferences. While AI model-based chatbots are more adaptive than rule-based chatbots, the training process of the AI model is different from the learning process of humans. Thus, it is hard to say that learners who learned by interacting with rule-based or AI model-based chatbots are engaged in an *interactive* mode of learning, even though the learners were engaged in *constructive* learning. The maximum level of cognitive engagement that can be achieved in learning primarily based on chatbot-learner interaction is *constructive*. However, previous studies have not focused on specific ways to induce a *constructive* mode of learning. To maximize the benefits of using educational chatbots, it is necessary to increase the cognitive engagement level to the *active* and *constructive* mode of learning, both of which are classified as active learning behaviors. Therefore, in this study, we will investigate whether the learning effect is different by different levels of cognitive engagement, specifically, *constructive* and *active* modes of learning.

For learning outcomes, the ICAP framework clarifies that *constructive* learning will lead to better learning achievement than active learning (Chi & Wylie, 2014). According to the ICAP framework, when students are engaged in the constructive mode of learning, they are more cognitively engaged in learning than in the active mode of learning. If a given learning activity is designed for constructive learning, students must infer and generate additional outputs using given learning materials and what they learned. Meanwhile, if a learning activity is designed for active learning, students can complete the tasks by manipulating learning materials and recalling what they learned (Chi & Wylie, 2014). Constructive (or active) chatbot learners are the learners who use chatbots that are designed to induce constructive (or active) modes of learning.

***H1a** – Compared to active chatbot learners, constructive chatbot learners will show a higher level of learning outcome than active chatbot users will.*

For learning motivation, the constructive mode of learning will lead to higher learning motivation than the active mode of learning. A higher level of cognitive engagement means learners are more challenged than in a lower cognitive engagement, leading to intriguing and motivating learners more (Blumenfeld, P., Kempler, T., & Krajcik, J., 2005). To the best of our knowledge, these findings have not been examined thoroughly in an educational context where educational chatbots are used as the main tools for learning. Our study will explore whether cognitive engagement levels during educational chatbot usage will positively impact learning outcomes and motivation. Here, the 'constructive chatbot' means the chatbot that leads a constructive mode of learning, whereas the 'active chatbot' means the chatbot that leads the active mode of learning. Therefore, our first hypotheses are as follows:

H1b *– Compared to active chatbot learners, constructive chatbot users will show a higher level of learning motivation than active chatbot users will.*

2.4.2. H2: Main Effect of Cognitive Determinants of Chatbot Humanness

Next, regarding computer-mediated communication (CMC) aspects, we will examine how the use of cognitive determinants of chatbot humanness impacts the learning experience. We manipulated six cognitive determinants of chatbot humanness, known to increase the user's perceived humanness. Such manipulation of chatbot humanness was grounded on previous literature (Seeger et al., 2021; Araujo, 2018, Rhim et al., 2022), where various determinants of humanness were incorporated to design humanlike chatbots. For example, Araujo (2018) incorporated the following elements to create a"huma'like disembodied chatbot: the use of informal language, the application of a human name, and initiating and concluding conversations through greetings. In Rhim et al. (2022)'s work, a chatbot designed for survey answering was humanized in four aspects: self-introduction, addressing respondents by name, utilizing adaptive response speed, and echoing respondents' answers. Survey respondents reported that the humanized chatbot was more humanlike compared to a non-humanized chatbot, which led to a higher social presence.

For learning outcome, it was shown that students who used a humanized educational interface agent were more successful than those who used a non-humanized one in terms of achievement retention of learning in science classes (Yılmaz and Kılıç-Çakmak 2012). Humanized (or non-humanized) chatbot learners are the learners who use chatbots that are humanized (or not humanized) using the cognitive determinants of perceived humanness. For learning motivation, it was reported that the humanized design of chatbots was positively correlated with learner motivation in voice-based chatbots (Ebadi & Amini, 2022). However, there has been less emphasis on how to Increase the humanness of chatbots in educational contexts. Additionally, there has been a lack of research to investigate whether the humanness of Ecs, especially text-based conversational agents, will affect the learning experience. Therefore, our second hypotheses are as follows:

H2a *– Compared to non-humanized chatbot learners, humanized chatbot learners will show a higher level of learning outcomes.*

H2b *– Compared to non-humanized chatbot learners, humanized chatbot learners will show a higher level of learning motivation.*

2.4.3. H3: Interaction Effect

In addition to the main effect of cognitive engagement in learning and cognitive determinants of humanness on the learning experience, we will examine whether these two independent variables have an interaction effect on learning. Compared to active activities, constructive activities induce a higher level of learning engagement. Thus, learners doing constructive activities will pay more attention to what chatbots say than learners doing active activities will. Such attention will enhance learner engagement in conversation with chatbots. Such engagement will be enhanced when using humanized chatbots rather than non-humanized chatbots (Sunny et al., 2021). Thus, we hypothesize that there will be an interaction effect between the cognitive engagement level and the cognitive determinants of chatbot humanness.

H3a *– When using constructive chatbots, humanized chatbot learners will show a higher level of learning outcome than non-humanized chatbot learners.*

H3b *– When using constructive chatbots, humanized chatbot learners will show a higher level of learning motivation than non-humanized chatbot learners.*

Chapter 3. Design of Chatbots

Chatbots developed for this study are designed to facilitate AI education in online learning environments. Their primary role is to guide users through a series of learning and problem-solving activities. To accomplish this, we create the chatbots using *DialogFlow*[①], a natural language understanding platform provided by Google that is well-suited for implementing simple rule-based chatbots. *DialogFlow* allows for the design and implementation of conversational user interfaces in various settings, including mobile and web applications, devices, bots, and interactive voice response systems. Conversational flows can be created using new 'Intents,' and user responses can be received in either multiple-choice or subjective forms using 'Custom events.' In addition, *DialogFlow* supports including multimedia content such as images and audio files, allowing for a richer and more engaging conversation.

The chatbots designed for this study guide learners, providing direction through 10 stages of learning and problem-solving activities. Detailed information regarding the learning and problem-solving activities is presented in *subsection 4.2. Procedure*. To begin a learning activity, the chatbot offers a button that leads the learner to a link where they can watch a *Youtube* video or train and test an AI model. Upon completing the given learning activity, the learner returns to the chatbot tab and clicks on a button to notify the chatbot that s/he finished the learning activity. In the case of non-humanized chatbots, problem-solving activities began immediately after the learning activity was completed. In contrast, in humanized chatbots, the chatbot

① https://cloud.google.com/dialogflow

engages in small talk with the user before initiating problem-solving activities. Once the problem-solving activity is completed, the chatbot guides the learner to the next learning activity.

In the present study, four versions of the chatbot are created to investigate the effects of two independent variables, namely the cognitive engagement mode of learning and the cognitive determinants of chatbot humanness, on learning outcomes/ motivation. For ease of reference, these chatbots are referred to as C-H, C-N, A-H, and A-N, where 'C,' 'A,' 'H,' and 'N' stand for 'Constructive,' 'Active,' 'Humanized' and 'Non-humanized,' respectively. Constructive (or active) chatbots refer to chatbots that are designed to induce constructive (or active) modes of learning. The distinction between the two modes of learning was made by four standards based on the ICAP framework (Chi & Wylie, 2014), which is further described in subsection 3.1. Humanized (or non-humanized) chatbots refer to chatbots that are (or are not) designed to be humanlike. The distinction between the humanized and non-humanized chatbots was made by manipulating six cognitive determinants of chatbot humanness based on the CMD framework (Seeger et al., 2021), explained in detail in subsection 3.2.

3.1. Manipulation of the Cognitive Engagement Mode of Learning: Constructive and Active

This subsection aims to provide a detailed description of the conditions under which constructive chatbots and active chatbots provide learning activities and problem-solving activities. The problems presented by constructive chatbots are designed to reflect one or more of the four characteristics of constructive learning.

In contrast, those presented by the active chatbots are designed to reflect one or more of the four characteristics of active learning. According to the ICAP framework, a constructive and active modes of learning is differentiated by a learner's overt behaviors that reflect her cognitive engagement level. In this study, the term 'constructive problem' refers to a problem intentionally designed to promote constructive learning. The following paragraphs provide a description of the four characteristics of constructive learning. However, it should be noted that the study does not explicitly measure whether learners learned constructively or actively when using constructive and active chatbots, respectively.

The first characteristic of constructive chatbots In the present study Involves providing learners with problems that require them to explain or summarize concepts in the video/ simulation in new sentences by reconstructing what was explicitly delivered. In contrast, active chatbots presented learners with problems that required them to either select an appropriate word to complete a given sentence or choose an appropriate sentence from a set of options. For example, after watching a short video that explained the basic definition of unsupervised learning, constructive chatbots asked, "Please explain what 'unsupervised learning' means by using the words *supervise* and *learning*." In doing so, learners have to not only recall what they learned in the video but also synthesize and apply the recalled knowledge in their responses. Meanwhile, active chatbots used precisely the same sentences that were already mentioned in the video, such as, "Please fill in the blanks of the following sentences: Unsupervised learning is one of the machine learning methods, where AI finds relations and patterns within the [label/data], without requiring humans to provide [result/answer]."

The rationale behind this approach is that constructing entire sentences requires a higher level of cognitive engagement and effort than simply filling in missing words or selecting from a given set of options. If provided with almost complete sentences, it is easier for learners to identify the correct answer by merely recalling relevant keywords or understanding the given sentences thoroughly. Here, deep understanding or further application of learning materials is not required. Thus, active chatbots require a lower level of cognitive engagement than constructive chatbots.

The second characteristic of the constructive chatbots In the present study involves providing learners with problems that require them to express their thoughts and reasons based on what they learned. In contrast, active chatbots provide argument and reason sentences with several blanks, and learners merely need to fill them in with short words. For instance, during problem-solving activities in the 4th stage, the constructive chatbots ask learners to identify whether the AI model trained during the preceding learning activity was supervised, unsupervised, or reinforcement learning and to explain their reasoning. In contrast, the active chatbots provide sentences with a few blanks that learners had to fill in to explain which of the methods used to train the AI model: supervised learning, unsupervised learning, and reinforcement learning. For example, learners had to fill in the blanks in the following sentence: “In this example, [supervised learning/ unsupervised learning/ reinforcement learning] was used, since this AI model was trained using data that have [blank], such as fish, other marine creatures, and trash.”

Thirdly, the problems provided by the constructive chatbots require learners to infer new information from the content learned in the preceding learning activity. In contrast, the problems provided by the active chatbots can be immediately solved by recalling the content learned in the preceding learning activities. For instance, after learning the basic knowledge about supervised and unsupervised learning in two videos, constructive chatbots asked, “How does the role of humans differ in two methods—supervised and unsupervised learning?” In contrast, active chatbots provide learners with complete sentences with few blanks where keywords are located, such as, “Unlike in supervised learning, humans do not need to provide [data/ labels of data] to AI.”

Fourthly, constructive chatbots require learners to apply the knowledge they have learned in a new situation not presented in the learning activity. In contrast, the active chatbots use examples already mentioned in the learning activity. Suppose that the video explained the fundamental concept of supervised learning and gives an example of how it is used by email services to distinguish between spam and non-spam emails. The constructive chatbots ask learners to describe, using specific words, how machine learning used by YouTube recommends a video to watch next. In contrast, the active chatbots provide a set of sentences that need to be arranged correctly to describe how machine learning classifies spam emails.

3.2. Manipulation of Cognitive Determinants of Chatbot Humanness: Humanized or Non-humanized

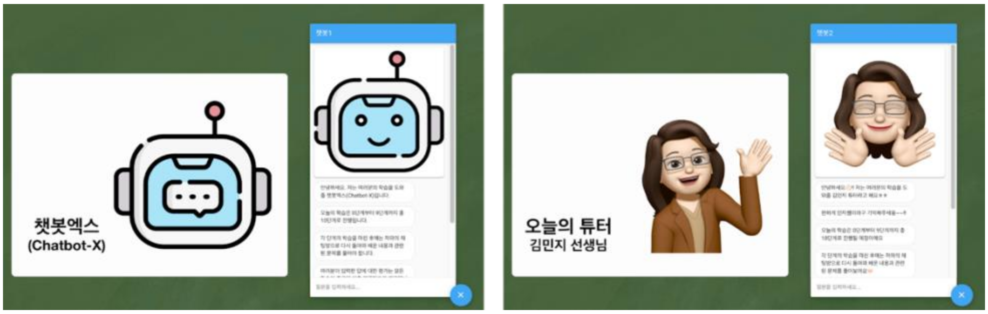

Figure 3. The user interface of non-humanized (left) and humanized (right) chatbots

This subsection provides a detailed description of the design and manipulation of humanized and non-humanized chatbots. Based on (Seeger et al., 2021)'s work, each of the three design dimensions (the first column of Table 1) can be expressed by several design elements. The six design elements reflected in each condition are described in the second column of Table 1, with the seven elements listed in the third and fourth columns for humanized chatbots and non-humanized chatbots, respectively. Each element in the non-humanized chatbot condition is the opposite of the corresponding element in the humanized chatbot condition. The implementation of each element is described in the following paragraphs respectively. Figure 3 depicts the user interface where learners interact with chatbots. The chatbots are operated through *DialogFlow Messenger* and are optimized for the Google Chrome environment.

Table 1. Design elements of humanized and non-humanized chatbots

Design dimension	**Design elements**	**Humanized chatbots**	**Non-humanized chatbots**
Identity of an agent	Humanlike visual representation	Anthropomorphic figure → Images of a human tutor character are represented.	Non-anthropomorphic figure → Images of a robot character are represented.
	Demographic information	Human names → The chatbot is named 'Minji Kim' and identified as a 'tutor.'	Non-human names → The chatbot is named 'Chatbot-X.'
Verbal expression	Social dialogue	Regular refreshment → The chatbot has a short casual talk before the problem-solving activity starts.	No refreshment → The chatbot does not talk to the learner other than through learning and problem-solving activities.
		Use of greetings → The chatbot starts and ends conversations when the learner says 'Hello~~'	No use of greetings → The chatbot starts and ends conversations when the learner says 'start' and 'end.'

		and 'Goodbye, Teacher Minji.'	
	Verbal style	Informal and casual expression → The chatbot uses 'Heyo-che' and gives grammatically incorrect variations at the end of verbs.	Formal expression → The chatbot uses 'Hao-che' and 'Hara-che.' and grammatically correct sentences.
Non-verbal expression	In-test emoji	Use of emojis within sentences	No use of emojis
	Stickers with facial expressions	Change in facial expressions → The tutor character shows various facial expressions and hand gestures and uses emojis within the text.	No change In facial expressions → The robot character has no nonverbal/paralinguistic expressions.

The first and the second design elements, (i.e., *picture of agents* and *name of agents*) are both related to the identity given to the chatbot. Specifically, the first element is manipulated by whether the agent's image is anthropomorphic or not. As shown in The second element is manipulated by whether the agent is named like a human or not. These design elements are incorporated based on previous studies conducted by (Araujo, 2018; Go & Sundar, 2019). In both the humanized and non-humanized chatbot conditions, an image and the name of the chatbot agent are presented on the left side of the webpage, while the messenger window is located on the right side.

The third design element (i.e., social dialogue) concerns the degree of interactivity between the chatbot and the learner. The difference in interactivity between humanized and non-humanized chatbots is based on previous research by Araujo (2018) that highlighted the importance of interaction in creating a more humanlike chatbot experience. The humanized chatbots have several short small talk sessions with learners, while the non-humanized chatbots do not. The humanized chatbots engage in interactions beyond guiding learners through learning activities, such as asking about learners' preferences or checking their memory of previous learning activities. These interactions are intended to enhance social presence and establish rapport between the chatbot and the learner. In contrast, non-humanized chatbots only guide learning activities and provide problems without engaging in any other forms of casual interaction.

Greetings, which is another way to have a social dialogue, pertain to the initiation and termination of interactions with the chatbot. The manipulation of this element is based on previous research conducted by Araujo (2018). In the case of humanized chatbots, the conversation is initiated when the user clicks on the utterance 'Hello~~', and it is terminated with the phrase 'Goodbye, tutor Minji~~.' This process is intended to mimic the way people start and end conversations in real life. In contrast, non-humanized chatbots initiate conversations when the user clicks on the utterance 'Start' and end them when the user clicks on 'End.' This approach not only differs from actual human communication but also gives users the impression of operating and shutting down a machine.

The fourth design element (i.e., verbal style) relates to the formality of chatbots' utterances. According to Araujo's study (2018), By implementing informal and casual linguistic features, chatbot users may perceive the chatbots as more humanlike. Thus, in the present study, humanized chatbots are designed to use informal expressions, specifically the 'Heyo-che' form, which is an informal and polite expression in the Korean grammar system (Chang, 1996). In contrast, non-humanized chatbots are designed to use formal and polite expressions, specifically the 'Hao-che' and 'Hapsho-che' forms (Chang, 1996). Additionally, humanized chatbots utilize causal variation commonly used by Koreans by modifying the vowel of the last word of the sentence or attaching certain suffixes such as ‘ㅂ’, ‘ㅇ’, or ‘ㅁ’. However, non-humanized chatbots are programmed to use only grammatically correct sentences without such linguistic modifications.

The fifth and sixth design elements (i.e., emoji and stickers with facial expressions) examined in this study pertain to the use of nonverbal and paralinguistic expressions in chatbot interactions. Humanized chatbots are designed to use facial expressions through tutor character images as well as emojis, a form of paralinguistic expression that can be delivered in text-based messengers. These emojis are primarily placed at the end of each sentence. In contrast, non-humanized chatbots do not use nonverbal or paralinguistic cues in their interactions. As shown in Figure 4, humanized chatbots employ various images, such as smiling faces and thumbs-up gestures, to convey positive emotions. However, in some situations with negative topics, the tutor character would show negative facial expressions, such as frowning. Non-humanized chatbots, on the other hand, use different types of expressions such as ‘...’, ‘?’, ‘!’, and bubbles on the robot character but do not utilize the nonverbal or paralinguistic expressions commonly found in computer-mediated communication.

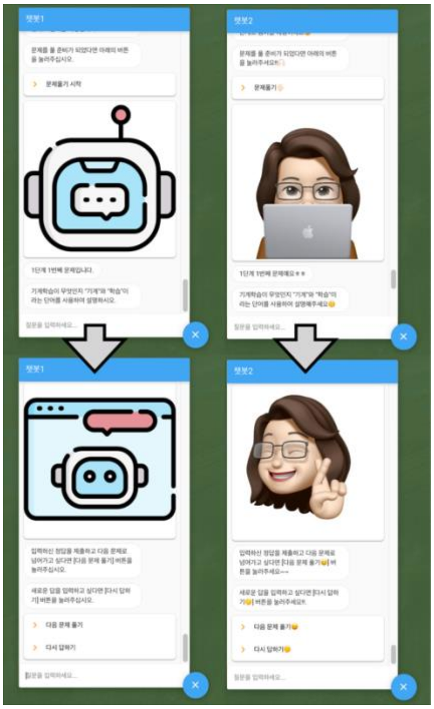

Figure 4. Comparison of a non-humanized chatbot (left) and a humanized chatbot (right)

Chapter 4. Method

4.1. Participant

The study recruited a total of 58 participants (12 male, 46 female) between the ages of 18 and 49 who possessed minimal or no prior knowledge of artificial intelligence. Participants were recruited through advertisements on second-hand trading platforms and online communities frequented by residents in Korea. Data collected from three participants were not included in the data analysis process, considering that they had internet connection issues or did not follow the instruction correctly during the experiment. As a result, data from a total of 55 participants were analyzed. Across the four conditions, there were no significant differences in age, gender ratio, pre-test score, and Affinity Technology Index score. The descriptive statistics were reported in Appendix 1.

4.2. Procedure

Prior to participating in the experiment, recruited participants were administered a pre-survey and pre-test. The pre-survey consisted of nine questions designed to assess participants' affinity for technology using the Affinity for Technology Index Scale (Franke et al., 2019). The pre-test involved participants solving ten questions, which were identical to those used in the post-test, to establish a baseline pre-test score. Both the pre-survey and the pre-test were computer-based and conducted via Google Forms. On the day of the experiment, participants accessed the virtual meeting room via a pre-notified Zoom link. The researcher randomly assigned each participant to four different experimental conditions.

The experiment followed a predetermined protocol, beginning with a 10-minute orientation session. In this introductory session, the researcher explained the procedures for accessing the chatbot website and for using a chatbot that guides learning and problem-solving activities. Participants were then instructed to access the chatbot website via Google Chrome. The website was developed to implement *DialogFlow* Messenger[②]. The researcher asked participants to share their Google Chrome page using the Zoom screen-share function. Each participant completed 10 sets of learning and problem-solving activities, with the entire session lasting approximately 40 to 50 minutes.

Following the experiment, participants were administered a post-survey consisting of 3 questions designed to measure the perceived humanness of the chatbot (MacDorman, 2006), 4 questions assessing perceived learning gains (Eom et al., 2006; Sun et al., 2008), and 25 questions measuring the Intrinsic Motivation Inventory. After the post-survey, the researcher provided participants with delayed feedback. The feedback was about incorrect responses during problem-solving activities. The post-test consisted of ten questions, identical to the pre-test questions. Participants were asked to complete this test within a ten-minute timeframe. Finally, a semi-structured interview was conducted to gather participants' opinions regarding their learning Ies and perceptions of how humanized the chatbot was.

Table 2. The entire session of learning and problem-solving activities

Stage	**Type of learning Activity**	**Specific task**

② https://cloud.google.com/dialogflow/es/docs/integrations/dialogflow-messenger

0	Learning	Video clip: What is AI? https://www.youtube.com/watch?v=Sxlz-r5FeCo&list=Pli9w7Ax2qVA9z-gT2YxmSyB0WJC3omJrX&index=1
	Problem-solving	Problem #0-1
1	Learning	Video clip: What is machine learning? https://youtu.be/IiyYsAMmmw4?list=Pli9w7Ax2qVA9z-gT2YxmSyB0WJC3omJrX
	Problem-solving	Problem #1-1, #1-2, #1-3
2	Learning	Exercise: Train an AI model to classify fish and trash https://studio.code.org/s/oceans/lessons/1/levels/2
	Problem-solving	Problem #2-1, #2-2, #2-3
3	Learning	Exercise: Test the AI model whether it can be used for cleaning the ocean https://studio.code.org/s/oceans/lessons/1/levels/3
	Problem-solving	Problem #3-1, #3-2
4	Learning	Exercise: Train an AI model to classify marine creatures and trash https://studio.code.org/s/oceans/lessons/1/levels/4
	Problem-solving	Problem #4-1, #4-2
5	Learning	Video clip: What is supervised learning? https://youtu.be/zXzFsWHToeg?list=Pli9w7Ax2qVA9z-gT2YxmSyB0WJC3omJrX
	Problem-solving	Problem #5-1, #5-2
6	Learning	Exercise: Train an AI model to classify fish by their shape or color https://studio.code.org/s/oceans/lessons/1/levels/6
	Problem-solving	Problem #6-1
7	Learning	Video clip: What is unsupervised learning? https://youtu.be/aJwUlyMsbX8?list=Pli9w7Ax2qVA9z-gT2YxmSyB0WJC3omJrX
	Problem-solving	Problem #7-1, #7-2, #7-3

8	Learning	Exercise: Train an AI model to classify fish by subjective criteria https://studio.code.org/s/oceans/lessons/1/levels/8
	Problem-solving	Problem #8-1
9	Learning	Video clip: What is AI ethics? https://youtu.be/rABDGSJm8tg?list=Pli9w7Ax2qVA9z-gT2YxmSyB0WJC3omJrX
	Problem-solving	Problem #9-1, #9-2

As described in Table 2, the learning activities and problem-solving activities completed by each participant were composed of 10 stages. Each stage typically lasted 4-5 minutes, with a total duration of 40-50 minutes for the entire session. The chatbot guided participants through alternating cycles of learning and problem-solving activities. There were two types of learning activities. The first one is to watch a 2-minute video clip of basic knowledge related to artificial intelligence. The video clips were developed by NAVER Connect Foundation[③], a non-profit organization for STEM education in South Korea. The first type of learning activities was conducted at the 0th, 1st, 5th, 7th, and 9th stages.

The second one Is to train and test AI models by supervised learning without programming, which is part of the AI for Ocean[④] program. 'AI for Ocean' is an AI education program offered by Code.org[⑤] and powered by AWS. Coge.org was founded to promote diversity and inclusivity in computer science education. Educational resources provided by this organization are designed to be accessible and engaging for students of all ages and backgrounds. The second type of learning

③ https://connect.or.kr/
④ https://code.org/oceans
⑤ https://code.org/

activities was conducted at the 2nd, 3rd, 4th, 6th, and 8th stages.

The problem-solving activities were designed to facilitate the recall and application of learning content covered in the preceding learning activity. Participants were presented with one to three problems at each stage. In contrast with active chatbots, constructive chatbots provide them with problems that require a constructive mode of learning. This mode of learning is cognitively more engaging than an active mode of learning, as it involves the application and synthesis of knowledge rather than mere recall. The entire set of the problems are described in Appendix 2.

4.3. Measurement

4.3.1. Post-experiment Survey

The dependent variables of the present study are learning outcome and learning motivation. Firstly, the measurement of learning outcome in the study was based on a pretest and a posttest consisting of nine identical questions each. The learning outcome score for each participant was computed by subtracting their pretest score from their posttest score. The tests were designed to assess participants' understanding of key concepts covered in the learning activities. Among the nine questions, three of them were considered difficult, requiring new knowledge and additional inference. Four of them were of medium difficulty that could be solved by applying what was learned in the learning activities. Two of them considered easy level that could be solved by simply recalling what was learned. The entire problems are described in Appendix 3.

Secondly, learning motivation was measured using a questionnaire adapted from Yin et al.'s study (Yin et al., 2021). The questionnaire was based on the Inner Motivation Index (IMI) originally developed by McAuley and colleagues (McAuley et al., 1989). This questionnaire measures intrinsic motivation in five dimensions: interest-enjoyment, tension-pressure, perceived choice, perceived competence, and perceived value. Each item was rated on a 7-point Likert scale, ranging from 1 (Strongly Disagree) to 7 (Strongly Agree).

To check the manipulation of the humanness level, a questionnaire that MacDorman (2006) developed to assess participants' perceptions of how the chatbot was humanlike was administered following the learning and problem-solving activities. Participants were asked to rate the chatbot on three scales: a nine-point mechanical versus humanlike scale, a nine-point strange versus familiar scale, and a ten-point eeriness scale. The scales ranged from very mechanical (1) to very humanlike (9), from very strange (1) to very familiar (9), and from not eerie (0) to extremely eerie (10).

The Affinity for Technology Interaction (ATI) Scale was also measured before each learner participated in the experiment. The ATI Scale questionnaire asks participants about their interaction with technical systems. The term "technical systems" refers to apps and other software applications, as well as entire digital devices (e.g., mobile phones, computers, TV, and car navigation). ATI can impact the chatbot using skill and, therefore, might impact the learning experience. Based on this, there should be no statistical difference in technology affinity between the four condition groups. Technology affinity was measured by the questionnaire developed by (Franke et al., 2019). The nine questions that comprise the questionnaire were rated on a 6-point Likert scale. The third, sixth, and eighth questions were reversely calculated.

4.3.2. Semi-structured Post-interview

Table 3 describes the questions that were asked to participants after the experiment was completed. Because the interview was semi-structured, additional questions were asked to elicit detailed explanations about responses further. For the ice-breaking, we asked participants the following questions. Next, to help participant brainstorm their learning through chatbot usage, questions about learning satisfaction and perceived learning gain were asked. To investigate what factors impacted the five dimensions of learning motivation, the five questions were asked.

Table 3. The semi-structured post-interview questions

Main topic	**Question**
Previous experience of chatbot usage, Usability	- Before participating in this experiment, have you heard about or personally used a 'chatbot' before? - Could you please let me know the difficulty level of the problems provided by the chatbot you used today? You can choose from high, medium, low, medium-high, or medium-low. - Did you experience any inconvenience in using the chatbot (in terms of functionality)?
Learning satisfaction	- Considering everything overall, could you please let me know your level of satisfaction with the chatbot problem-solving activity? You can choose from high, medium, low, medium-high, or medium-low.
Perceived learning outcomes	- Did the activity of "solving problems with a chatbot" help you acquire the learned content? Or did it not particularly help you in acquiring the content?
Learning motivation	[Interest-enjoyment] - Do you think solving problems with a chatbot was helpful in making the learning of artificial intelligence (AI) enjoyable, or do you think it's not necessarily related to enjoyment? [Tension-pressure] - Did you feel any pressure or tension while solving problems with a chatbot, or did you not particularly feel that way? [Perceived choice] - Do you have any intention of using the method of solving problems with a chatbot in future learning situations, or do you not have any specific intention? Please feel free to share

	your opinion. [Perceived competence] - From your own perspective, do you consider yourself proficient in solving problems using a chatbot, or do you think you need more time to become proficient? [Perceived value] - Please answer with a yes or no: Do you consider a chatbot to be a useful learning tool? If possible, please provide reasons for your answer.
Humanness of the chatbot	- Do you remember the name of the chatbot you used in today's experiment? - Could you please describe your "impression" or "feeling" about the chatbot using 1-2 words? It can be a noun, adjective, adverb, or any other part of speech. - Was the chatbot's tone more mechanical or human-like? - In what aspects did it feel more mechanical, and in what aspects did it feel more human-like? *For humanized chatbot users* - What factors can contribute to making the chatbot feel more human-like? 1. Human-like avatar or visual representation. 2. Using a human name. 3. Informal or casual tone of speech. 4. Human facial expressions and gestures. 5. Engaging in conversation in a more interactive manner. - Which one do you think is more helpful for learning: humanlike or machinelike manner?

Finally, to explore whether participants perceived the chatbot as humanized or non-humanized and what impacted such perception, two questions were asked: 'Could you explain the image of the chatbot you used in one or two nouns, adjectives or adverbs?' 'How did you perceive the chatbot's dialogues: either humanlike or machinelike?'. Also, additional questions, such as 'What is the main component of the chatbot that made you perceive"the 'hatbot as machinelike?' were asked according to the participant's responses.

4.3.3. Response Data Collection

Beyond the utilization of self-report measures, including questionnaires and interviews, we also collected response data that were generated during problem-solving. The collected response data was labeled to answer our research question 1—whether constructive (or, active) chatbots successfully induce a constructive (or, active) mode of learning.

4.4. Data Analysis

To check the manipulations and test the main/ interaction effects of the two independent variables, we used mixed approach for data analysis where both quantitative and qualitative data were analyzed. Table 4 summarizes the specific data analysis methods that were employed for each goal. Detailed process of quantitative and qualitative data analyses will be described in subsections 4.4.1 and 4.4.2, respectively.

Table 4. The specific data analysis methods and their goals

Goal	Quantitative data analysis	Qualitative data analysis
Manipulation check of cognitive engagement mode of learning	-	Response data analysis
Manipulation check of cognitive determinants of chatbot humanness	Welch's two-sample t-test	Interview data analysis
H1, H2, H3	Two-way ANOVA, Aligned Rank Transform for Nonparametric Factorial ANOVAs	Interview data analysis

4.4.1. Quantitative Data Analysis

For manipulation check of using cognitive determinants of chatbot humanness, we used a two-sample t-test to check whether humanized chatbot users actually perceived their chatbots as more humanized than non-humanized chatbot users did. Before this statistical testing, we standardized the variables so that the value should be within the 0-1 range.

For testing H1, 2, and 3, we used two-way ANOVA. The assumption tests for two-way ANOVA were satisfied. Firstly, the assumption for normal residuals was examined through the Shapiro-Wilk test. The results confirmed that the residuals are normally distributed for most of the hypotheses, with the p-value not being less than the significance level of 0.05. In addition, the assumption for an equal variance for Levene's test was satisfied for all the hypotheses, with the p-value not being less than the significance level of 0.05.

Lastly, the assumption for independence for the Durbin-Watson test was satisfied for learning outcome and the second to fifth dimensions of learning motivation, with the DW statistic being approximately between 1.5 and 2.5. In detail, for the first dimension of learning motivation (IMI_1), the DW statistic was 2.7476, indicating there is little concern for negative autocorrelation. In Appendix 4, the statistics of the assumption tests were described.

Exceptionally, for the second dimension of learning motivation (i.e., tension-pressure), we used the Aligned Rank Transform Factorial ANOVA test (Wobbrock et al., 2011) for testing H1b, H2b, and H3b. This is because the normal distribution assumption was not satisfied for the second dimension of learning motivation. Specifically, the *p-value* of Levene's test was 0.003**. However, considering that the other two assumptions were satisfied, we used the Aligned Rank Transform

Factorial ANOVA test, a substitute for the two-way ANOVA test for non-parametric variables (Wobbrock et al., 2011). Specifically, the dependent variable was aligned and ranked to be transformed into a new value (rank). The newly computed value is used for the two-way ANOVA test.

4.4.2. Qualitative Data Analysis

In addition to quantitative data analysis, we also analyzed qualitative data to investigate further the impact of cognitive engagement modes of learning and cognitive determinants of chatbot humanness on learning outcomes and motivation. Two types of qualitative data were analyzed: (1) response data during problem-solving and (2) post-interview data.

(1) Response data during problem-solving

Responses made by users during problem-solving were collected. For manipulation check of cognitive engagement mode, we labeled each response as whether it can be classified as an output of constructive (or, active) mode of learning. The ICAP framework assumes that cognitive engagement is evaluated through overt behavior and externalized output. It is ambiguous and challenging for an external researcher or instructor to judge how internally cognitively engaged a learner is. For that reason, in this study, the manipulation check to verify whether the intended modes of cognitive engagement in learning were induced was done by using response data. Response data are naturally generated during the problem-solving process, and thus can reflect more precise evidence of learners' cognitive engagement than self-reported measurements can.

For the responses made by constructive chatbot users, we labeled each response as 'constructive' when a given response met at least one of the four standards of the constructive mode of learning mentioned in subsection 3.1. This labeling approach was taken due to the possibility that constructive chatbot learners gave responses by active modes or passive modes of learning, even though the constructive chatbots were designed to induce constructive modes of learning. For instance, the chatbot asked a learner to give her opinion and justify it using what she learned. When the learner's response only includes the main insistent without grounds, the response was not classified as an externalized output of constructive learning. This is because it violated the second characteristic that was employed to induce constructive modes of learning. Then, for each learner, we calculated the '*constructive* score' as the ratio of the number of problems labeled as constructive to the entire number of problems given.

For the responses made by active chatbot users, we also labeled each response as active when a given response was not made from random answering or a mistake. Random answering was detected by researchers, and then the researcher asked the learner whether it was randomly answered or not right after the post-feedback was given. Mistakes were observed when a learner made an answer without reading the conditions or the number of blanks. Then, for each learner, we calculated the '*active* score' as the ratio of the number of problems labeled as constructive to the entire number of problems given.

(2) Post-interview data

Data from the post-interview was analyzed using a combination of inductive and deductive coding methods (Corbin & Strauss, 2015). This was because participants often gave answers not directly related to the question but somewhat related to another topic, although the interview questions were semi-structured. For example, when a learner was asked, 'Did this chatbot make you enjoy learning?'

some learners mentioned that 'Yes, the chatbot was useful for obtaining knowledge.' Thus, we first categorized the transcripts by sentence, open-coded each sentence, and grouped the codes into themes.

Firstly, based on the structural coding method, the three coders independently categorized each of the sentences of transcripts. The categories pertain either to learning effects or the humanness of chatbots. Secondly, using an inductive approach, the coders open-coded all the categorized sentences line-by-line and created initial codes. Thirdly, similar codes were merged to form themes, generating a total of 25 themes. The second a"d th'rd steps were iteratively conducted to meet a sufficient level of agreement between the three coders. A satisfactory inter-rater reliability score was achieved, with the Fleiss' kappa score of 0.71.

● ***Labeling rules for comments that relate to the learning effect***

Within the array of sentences derived from the interview script data, those pertinent to learning were segregated into two parts: 'cause' and 'result.' The 'cause' part is a factor that influences learners' learning experience. Conversely, the 'result' part refers to the consequence of such causes on the learners. For example, if a learner reported that 'I could enjoy learning because the refreshment was interesting and fun,' the 'cause' part is 'refreshment' and the 'result' part is 'learning enjoyment.'

As demonstrated in Table 5 (a), the first categorical label within the 'result' part signifies either a positive or negative outcome. The second categorical label encapsulates three aspects: 1) emotional benefits: User perceptions about the learning process being (or not being) enjoyable, entertaining, or boring, 2) practical benefits: user viewpoints about the process being (or not being) beneficial for learning or acquiring knowledge, and 3) general expressions: generic feedback such as good, great, bad, or not good.

If a given sentence during an interview was labeled as a positive result, the 'cause' part was coded by the labels described in Table 5 (b). The first categorical labels are as follows: 'problem-related,' 'curriculum-related,' 'dialogical flow-related,' 'design element-related,' and 'social presence-related.' For example, if a learner said, "As I just need to follow the direction of the chatbot, it was really useful for reminding me what I learned." the 'result' part is labeled as positive. Thus, the 'cause' part is labeled as 'curriculum-related' and 'the curriculum is already set.'

If a given sentence during an interview was labeled as a negative result, the 'cause' part was coded by the labels described in Table 5 I. The first categorical labels are as follows: 'problem-related,' 'feedback-related,' 'dialogical flow-related,' 'design element-related,' and 'structure-related.' For example, if a learner said, "As I needed to write my own opinion, it was not good." the 'result' part is labeled as positive. Thus, the 'cause' part is labeled as 'curriculum-related' and 'the curriculum is already set.'

Other than the two independent variables (i.e., cognitive engagement level and cognitive determinants of chatbot humanness), the labels used for the 'cause' part included factors that were controlled over the four experimental conditions. For instance, 'instant reply back', more specifically, 'the chatbot reacted immediately after I finished answering,' pertained to both constructive and active conditions, and both humanized and non- humanized conditions.

Table 5. Labels used for open coding the post-interview transcript data

(a) Labels used for coding the 'result' part of each sentence during the interview

Category	**Theme**	**Label**
Positive	Emotional benefits	Enjoyable, interesting, fun, not burdensome

result	Practical benefits	Acquire new knowledge, review/remind/memorize learning materials, motivate to continue learning
	General expressions	Good, satisfying, okay
Negative result	Emotional benefits	boring, burdensome, not interesting, not enjoyable
	Practical benefits	Not useful/helpful, nothing more than a short review
	General expressions	Not good, unsatisfactory, not okay

(b) Labels used for coding the ‘cause’ part of each sentence during the interview (with a positive result)

Category	**Theme**	**Label**
Cause of a positive result	Problem-related	Questions requiring critical thinking skills.
		Addressing the core concepts of the learning material through questions
	Design-related	Having the conversation history visible in the chat room
		Having readability/ intuitiveness/ ease of use
		Using characters or emoticons
	Curriculum-related	Providing a predefined curriculum/ step-by-step learning guidance
		Serving as a prompt for review and recall
		Creating a sense of engagement, similar to games or quizzes
		Considered satisfactory compared to other learning methods
	Conversation progression-related	Allowing learners to progress at their desired pace
		Providing immediate responses
		Allowing for answer modifications
		Offering non-learning related conversations
	Social presence-related	Reducing pressure by interacting with a non-human entity
		Giving the feeling of being accompanied by someone
		Fostering a sense of interaction

(c) Labels used for coding the 'cause' part of each sentence during the interview (with a negative result)

Category	Theme	Label
Cause of a negative result	Feedback-related	Lack of immediate feedback
		Limited flexibility in feedback
	Problem-related	Difficulty in providing descriptive or subjective answers
		Challenges in understanding the intention of questions/ High difficulty level of the questions
		Inclusion of hints within the problem
		Requiring specific conditions or constraints when providing answers
		Concerns about making mistakes/ feeling burdened during problem-solving
	Design-element related	Poor visibility/ excessive chat content
		Perceived stiffness or static nature of the interaction
	Structural limitation	Difficulty in anticipating following content
		Inappropriateness for learning advanced or complex topics
		Inability to ask questions or engage in conversation, focusing solely on problem-solving

● ***Labeling rules for comments that relate to chatbots' humanness***

Concerning speech style, we asked both humanized and non-humanized chatbot users about whether they perceived their chatbots as machinelike or humanlike. According to their response, we asked what specific factor impacted their perception, and factors were labeled into three main codes: Dialog-related factors, sentence-related factors, and others. Table 6 (a) shows the factors that made non-humanized chatbot users perceived their chatbots as machinelike (upper part) and humanlike (lower part). Table 6 (b) shows the factors that made humanized chatbot users perceived their chatbots as humanlike (upper part) and machinelike (lower part).

Table 6. Labels used for coding 'cause' part of why the learners perceived the speech style of their chatbots either as machinelike or humanlike

(a) Non-humanized chatbot users (C-N, A-N conditions)

Category	Theme	Label
Why non-humanized chatbot users perceived a chatbot's speech style as machinelike	Speech-related	the repetitive pattern of conversation or word usage/ consistent speech speed
		Lack of emotions, reactions, and everyday conversations
	Sentence-related	Lengthy sentences
		Perceived as formal or written language
		Short and concise answers
	Others	Presence of a structured progression
		Feeling of responding to a survey
		Overall impression
		Related to post-feedback
		Providing images of robot
Why non-humanized chatbot users perceived a chatbot's speech style as humanlike	Speech-related	Diverse speech patterns
		Immediate responses
		Preconception that chatbots talk based on human-generated inputs
		Feeling of engaging in a conversation
	Sentence-related	Lack of awkwardness
		Use of colloquial or informal language
		Avoidance of complex vocabulary
		Not using lengthy sentences
	Others	Provided questions
		Presence of typos

(b) Humanized chatbot users (CH, AH conditions)

Category	Theme	Label

<table>
<tr><td rowspan="5">Why humanized chatbot users perceived a chatbot’s speech style as machinelike</td><td rowspan="2">Speech-related</td><td>Providing anticipated responses only</td></tr>
<tr><td>Recognizing input as pre-existing</td></tr>
<tr><td rowspan="2">Sentence-related</td><td>Using imperative tone</td></tr>
<tr><td>Using mechanical language</td></tr>
<tr><td colspan="2">Others</td></tr>
<tr><td rowspan="5">Why humanized chatbot users perceived a chatbot’s speech style as humanlike</td><td rowspan="2">Speech-related</td><td>Providing non-learning related responses</td></tr>
<tr><td>Providing human-generated responses</td></tr>
<tr><td rowspan="2">Sentence-related</td><td>Using colloquial language</td></tr>
<tr><td>Using emoticons or symbols</td></tr>
<tr><td colspan="2">Others</td></tr>
</table>

Chapter 5. Result

5.1. Manipulation Check

5.1.1. Cognitive Engagement Modes of Learning

As explained in subsection 4.4.2, the constructive [active] score was defined and computed as the ratio of the number of problems labeled as constructive [active] to the entire number of problems given. We were able to manipulate the cognitive engagement level in learning by designing two versions of chatbots that induced constructive and active modes of learning, respectively. The average constructive score was 0.89, proving that learners solved 90% of the questions by constructive mode of learning. Out of a total of 29 *constructive* chatbot learners, 28 of them scored greater than 0.75 (0-1 scale). This indicates that the majority of constructive chatbot learners (96.6%) were cognitively engaged in constructive modes of learning for three-fourths of their problem-solving. The average *active* score was 0.91, showing that learners solved 90 % of the questions by active mode of learning. Out of 26 *active* chatbot learners, 25 of them scored greater than 0.84.

5.1.2. Cognitive Determinants of Chatbot Humanness

(1) Quantitative Data Analysis

As a result of the t-test, we confirmed that the humanized chatbot users perceived their chatbots as more humanlike than the non-humanized chatbot users did (t = 1.696, df = 48.19, $p\text{-}value$ = 0.048*). Specifically, humanized chatbot learners (m = 0.208, sd = 1.158, n = 29) perceived a higher level of humanness than non-humanized chatbot learners (m = -0.232, sd = 0.742, n = 26). The standardized score of perceived humanness was used since two questions measuring perceived humanness were a 9-point Likert scale while one question was a 10-point Likert scale. Table 7 and Figure 5 show how the standardized score of perceived humanness differs using the cognitive determinants of chatbot humanness. The mean scores of perceived humanness of humanized chatbot learners (n=29) and non-humanized chatbot learners (n=26) were 0.208 and -0.232, respectively.

Table 7. The perceived humanness scores of two groups: humanized chatbot users and non-humanized chatbot users

Standardized score of perceived humanness	Sample size	Mean	Standard deviation
Humanized chatbot users	29	0.208	1.158
Non-humanized chatbot users	26	-0.232	0.742

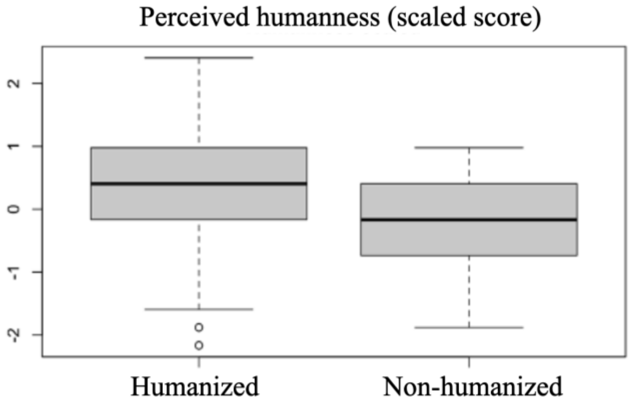

Figure 5. Boxplots of the perceived humanness scores of two groups: humanized chatbot users and non-humanized chatbot users

(2) Qualitative Data Analysis

● ***The impression of chatbots***

Regarding the impression of the chatbot and the perceived humanness of speech style, the qualitative analysis of learners' responses supported the quantitative analysis results. Firstly, it was confirmed that the cognitive determinants of chatbot humanness significantly influenced the impression of the chatbot (i.e., whether the chatbot was perceived as closer to a human or machine). Approximately 41% (n = 12) of humanized chatbot learners said that the chatbot felt like a teacher, an educator, or an instructor. Around 28% (n = 8) mentioned that the chatbot was kind, friendly, or funny. While these adjectives can also be used to describe machines, they are generally high-warmth expressions known to be associated with feeling humanness (Roy & Naidoo, 2021). On the other hand, the majority of non-humanized chatbot learners, about 77 % (n = 20), reported that they perceived the chatbot as a machine, a program, a robot, or an AI.

● ***Speech style of humanized chatbots***

Approximately 72 % (n = 21) of humanized chatbot users reported that their chatbots talked in a humanlike way. 9 of them (CH_1, 3, 6, 8, AH_3, 5, 6, 10, 11) mentioned that chatbots' giving utterances for refreshing made them perceive chatbots as humanlike. 7 of them (CH_4, 6, 10, 13, AH_1, 8, 12) attributed to the use of conversational style, and 3 (CH_7, AH_7, 12) mentioned the use of emojis and symbols led to perceiving humanness.

> *"When I saw the chatbot saying something like "The teacher also enjoys practicing that way," it made me feel that it's not just throwing questions but also engaging in conversation. Of course, it could be a prepared scenario, but it gave the impression of being more human-like." (CH_6)*
>
> *"Now that we are familiar with messengers, I often thought that the chatbot's conversational style was similar to chatting with friends on a messenger. And even on online learning platforms, when leaving questions or receiving answers, sometimes they respond in a conversational tone. With that similar*

experience, I felt that it had less of a robotic feel." (CH_4)

● ***Speech style of non-humanized chatbots***

About 81 % (n = 21) of non-humanized chatbot users reported that the chatbot spoke like a machine. Regarding specific factors that let users perceive such a low level of humanness, 7 of them (CN_4, 6, 12, 13, AN_1, 2, 9) mentioned the repetitive flow of dialogue and the use of similar words, 5 of them (CN_3, 7, 10, 14, AN_6) attributed to the sentences with written style, and 3 of them (CN_12, AN_8, 14) mentioned lack of emotional expressions, reactions, and small talks. They responded that the chatbot would feel more humanlike if it used social dialogues and informal verbal styles, which were utilized in humanized chatbots. Such a suggestion indicated the demand for other factors that can increase the humanness of chatbot, such as motivational comments and adaptive feedback. However, considering that such factors could have been strong compounding variables in our study, our chatbots did not provide learners with such factors. If the chatbots designed in our research experiment were to be commercialized and to target users from diverse backgrounds, it would be beneficial to consider incorporating more diverse types of cognitive determinants of chatbot humanness.

Additionally, to the non-humanized chatbot users who responded that they perceived the chatbot as machinelike, we asked for suggestions to design a chatbot more humanlike. Concerning the flow of dialogue, 4 users (CN_5, 13, AN_9, 13) suggested 'using refreshing talks, emotional reactions, supportive messages,' and 3 users (CN_4, 11, AM_4) suggested 'providing sentences differently by user's input.' Regarding the sentence style, 5 users (CN_3, 10, AN_2, 4, 9) recommended using informal and conversational expressions.

"For example, when someone provides encouragement or support during

moments of struggling or stumbling, it creates a sense of home that only a person can provide. If such elements are added, it would give the impression of having more human-like conversations." (CN_5)

"I believe that speech style is quite important. If the chatbot uses a friendly tone rather than a rigid one, incorporates occasional humor or human-like elements, even a simple chatbot can feel more approachable." (AN_9)

"I think using colloquial language or informal speech can make it appear more human-like. It may give a more personable impression." (CN_3)

Interestingly, even though the quantitative analysis showed that humanized chatbot users perceived their chatbot as more humanlike than the non-humanized chatbot users did, 13 non-humanized chatbot users reported that the chatbot spoke like a human. They attributed to various factors, such as 'the use of informal and conversational sentences (CN _7, 9)' and 'well-structured sentences (CN_4, 15, AN_12).'

5.2. H1: Main Effect of Cognitive Engagement Mode of Learning

5.2.1. Quantitative Data Analysis

As hypothesized in H1a, there was a significant main effect of cognitive engagement modes on learning outcomes ($F(1, 51) = 8.821$, $p = 0.00453$**, $\eta_p^2 = 0.13$). However, unlike hypothesized in H1b, no significant main effect on learning motivation was observed. Table 8 shows the two-way ANOVA test results on the main effect of cognitive engagement modes. For the second dimension of learning motivation (i.e., tension-pressure), the result of ART ANOVA was reported.

Table 8. Two-way ANOVA and ART ANOVA tests to examine the main effect of cognitive engagement on learning outcomes and motivation (C and A refer to 'Constructive' and 'Active' chatbot learners, respectively.)

Dependent variables		Condition	Mean	Standard deviation	F-value $F(1, 51)$	p-value	Effect size
Learning outcome		C	5.42	1.85	8.82	0.004**	0.15
		A	3.89	2.12			
Learning motivation	1	C	5.84	0.94	3.48	0.07	0.06
		A	5.37	0.92			
	2	C	2.44	1.28	0.03	0.87	0.00
		A	2.58	1.36			
	3	C	4.35	0.79	2.07	0.16	0.04
		A	4.68	0.87			
	4	C	5.14	1.05	0.01	0.91	2.76e-04
		A	5.11	0.99			
	5	H	5.75	1.00	0.05	0.83	9.09e-04
		N	5.81	0.90			

$p^{**} < 0.01$

5.2.2. Qualitative Data Analysis

● ***Did Cognitive Engagement Mode of Learning Impact Learning Outcomes?***

In accordance with the quantitative analysis, we found evidence in favor of the effect of cognitive engagement modes on learning gain in the post-interview. 4 out of 13 CN chatbot users (CN_8, 9, 10, 13) and 6 out of 15 CH chatbot users (CH_3, 5, 6, 7, 10, 13) reported that the problems required a deep understanding of what they learned and made them think further what they learned from learning materials. In contrast, only one AM chatbot user (AN_2) and one AH chatbot user (AH_11) reported that the problem intrigued deep level of the dimensions. AM_3, 4, AH_13

"Since the problems seemed a bit difficult, I found myself more focused and engaged in solving them, which allowed me to acquire more of the learned content." (CN_10)

"Now that the questions are structured to require more critical thinking, I think such question composition is more suitable for learning." (CH_3)

Through the observation of chatbot usage, we found indirect evidence supporting that the high cognitive engagement level that learning activity induces leads to better learning outcomes. 9 Active chatbot users tended to make a mistake at the end, and they admitted that it was their mistake because they started to lose concentration as they got too used to clicking the answer. Specifically, the last question was as follows: *We need to examine various situations that can arise due to artificial intelligence and collect opinions from [people from various fields/multiple developers] to set standards. This is because if artificial intelligence only learns from the choices of a minority, it cannot make an [objective/biased] judgment.* The answers were 'people from various fields' and 'objective', because the lecture delivered that if artificial intelligence only learns from the choices of a few individuals, it can lead to 'biased' decisions. This question was formulated as a negative version of that statement. While the researcher gave delayed feedback, the active chatbot learners also reported that they lost concentration and did not read the last question as thoroughly as they did in the early problem-solving stage. However, among the constructive chatbot users, there were hardly any who made mistakes as the session progressed.

● ***Did Cognitive Engagement Mode of Learning Impact Learning Motivation?***

In accordance with the quantitative analysis results, we could find a few pieces of evidence in favor of why the main effect of cognitive engagement level was not significant on learning motivation. Five constructive chatbot users (CN_8, 10, 11, CH_9, 10, 15) reported that constructive questions were too difficult, which decreased learning satisfaction or enjoyment. On the other hand, four constructive chatbot users (CN_8, 13, CH_3, 13) commented that constructive questions were intriguing and interesting since such questions made them think to reorganize what they learned.

> *"The problem-solving didn't aid in my learning. This is because the information asked in the problem was slightly more difficult than what I had learned, which made me feel discomfort." (CN_11)*
>
> *"I found many problems that required me to think about applications to other examples. The fact that I had to think deeply when typing answers made the problem-solving process interesting." (CH_13)*

Similarly, comments on both positive and negative impacts of an active mode of learning were reported by active chatbot users. In other words, we could not find evidence that active mode of cognitive engagement led to differences in learning motivation.

5.3. H2: Main Effect of Cognitive Determinants of Chatbot Humanness

5.3.1. Quantitative Analysis Result

As hypothesized in H2a, there was a significant main effect of cognitive engagement level on learning outcome ($F(1, 51) = 4.37$, $p = 0.04$*, $\eta_p^2 = 0.08$). However, unlike hypothesized in H2b, no significant main effect on learning motivation was detected. Table 9 shows the two-way ANOVA test results on the main effect of cognitive engagement level. For the second dimension of learning motivation, the result of the aligned ranks transformation ANOVA (ART ANOVA) was reported.

Table 9. Two-way ANOVA and ART ANOVA test results to examine the main effect of chatbot's humanness on learning outcomes and motivation (H and N refer to 'humanized' and 'non-humanized' chatbot learners, respectively.)

0.04*		**Condition**	**Mean**	**Standard deviation**	**F-value** $F(1, 51)$	**p-value**	**Effect size**
Learning outcomes		H	4.17	2.28	4.37		0.08
		N	5.23	1.80			
Learning motivation	1	H	5.58	0.96	0.05	0.82	1.05e-03
		N	5.63	0.97			
	2	H	2.34	1.27	2.33	0.13	0.04
		N	2.69	1.35			
	3	H	4.43	0.89	0.49	0.49	9.49e-03
		N	4.60	0.79			
	4	H	5.08	1.12	0.13	0.73	2.44e-03
		N	5.18	0.89			
	5	H	5.68	1.00	0.62	0.44	0.01
		N	5.88	0.88			

$p^* < 0.05$

5.3.2. Qualitative Analysis Result

● ***Did the cognitive determinants of chatbot humanness impact learning outcomes?***

We could find evidence that might explain why cognitive determinants of chatbot humanness brought about an unexpected impact on learning outcomes. Both humanized (CN_2, 3, 5, 9, 12, 13 & AN_3, 5) and non-humanized chatbot learners (CH_8, 10, 14 & AH_3, 5) reported that they benefited from the fact that chatbots led the entire learning activities, so that learners were required just to follow the chatbots' guidance.

> *"The repeating structure of the chatbot presenting problems and me solving them was well designed. I was able to learn step-by-step, which I think led to rapid learning effects." (AH_5)*
>
> *"I liked the fact that I could learn systematically with a defined curriculum, which was better than studying alone." (CH_14)*
>
> *"Human teachers can be less systematic than chatbots, in my opinion. For example, people may simply categorize stages as basic and intermediate levels. However, with the chatbot, it felt like it was teaching me slowly from step 0 to 9, which I felt was helpful for learning." (AN_05)*
>
> *"I kept having the feeling that a clear step-by-step learning curriculum was being provided and I was being guided, almost like a roadmap was being drawn out for me." (CN_3, 13)*

In addition, unexpected comments were reported from non-humanized chatbot learners. Both humanized (AH_1, 6, 10, 11) and non-humanized chatbot learners (CN_1, 7 & AN_6) mentioned that the feeling of interacting with the chatbot contributed to obtaining knowledge and willingness to learn.

> *"Just reading the content alone can be somewhat boring and monotonous, but when I used the chatbot, I was able to have a conversation with Teacher Minji and immediately confirm information. Learning in this way felt much more useful and beneficial." (AH_10)*
>
> *"Since we were learning through interactive conversation, it didn't feel like studying alone but rather like being part of a study group. (...) It felt like it was helpful for my learning." (AN_6)*

● ***Did the cognitive determinants of chatbot humanness impact learning motivation?***

We could find a few pieces of evidence that might explain why learning motivation was not significantly impacted by the cognitive determinants of chatbot humanness. Both humanized chatbot learners (CH_8, 9, 10, 12 & AH_1, 6, 9, 10, 11) and non-humanized chatbot learners (CN_7 & AN_1, 6, 11) reported that they felt their interaction with the chatbots as humanlike. Also, four learners (CN_5, 7 & CH_5, 9) commented that they felt as if they existed with a human being during problem-solving with the chatbots.

> *"I felt a bit more emotionally comfortable because I felt as if I was having a natural conversation situation." (CH_8)*
>
> *"I felt like I was receiving some kind of face-to-face guidance, which made studying less boring." (CH_9)*
>
> *"I think the fact that the chatbot attempted to have a conversation like a human, thereby making the conversation more natural, had a positive influence in making problem-solving less boring and more interesting." (AH_10)*
>
> *"Just as you would concentrate if there were someone who keeps asking you questions, it seems like the chatbot was playing that role." (CN_7)*
>
> *"It didn't feel like studying alone, but more like studying with other mates. Having the chance to choose the right answer while exchanging messages back and forth seemed to be helpful." (AN_11)*

Despite the fact that H2b was not satisfied, one of the cognitive determinants of perceived humanness, which is the use of visual representation of humanlike characters, might have impacted learning motivation. Four (AH_1, 6, 8, 13) reported that the use of images of *Teacher Minji Kim* and emojis positively impacted learning enjoyment.

> *"It seemed more intriguing when a character that looked like a teacher or something, appeared and asked questions, which made me feel those questions less difficult, despite some questions were hard to me." (AH_13)*

5.4. H3: Interaction Effect

5.4.1. Quantitative Data Analysis

No interaction effect was observed on learning outcomes, which was in contrast with H3a, but there was a significant interaction effect between cognitive engagement mode of learning and cognitive the cognitive determinants of chatbot humanness on learning motivation. However, such significant interaction effects were disordinal, while an ordinal interaction effect was anticipated in H3b. Table 10 shows the results of two-way ANOVA tests and the Aligned Rank Transform ANOVA on the interaction effect between cognitive engagement level and cognitive determinants of chatbot humanness on the five dimensions of learning motivation.

Figure 6 shows significant interaction effects that were reported for the second and fourth dimensions of learning motivation. H3b was tested as true for the second dimension of learning motivation (i.e., tension-pressure). When using constructive chatbots, cognitive determinants of humanness gave a positive effect on tension-pressure; in contrast, when using active chatbots, cognitive determinants of humanness gave a negative effect on tension-pressure. For the fourth dimension of learning motivation (i.e., perceived competence), a significant interaction effect was reported but directly opposed to H3b. In other words, when using constructive chatbots, cognitive determinants of humanness gave a negative effect on perceived competence; in contrast, when using active chatbots, cognitive determinants of chatbot humanness gave a positive effect on perceived competence.

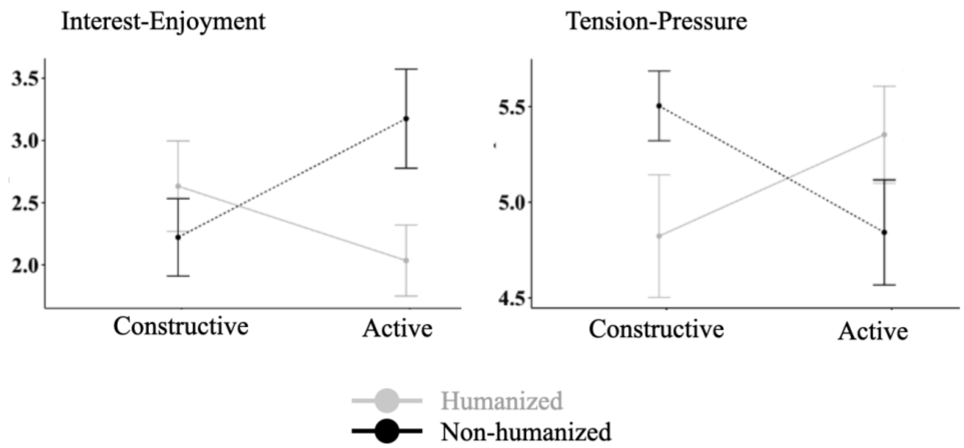

Figure 6. Interaction effect between cognitive engagement mode of learning and cognitive determinants of chatbot humanness on perceived competence (the second and the fourth dimensions of learning motivation)

Table 10. ANOVA result of testing the interaction effect between cognitive engagement mode of learning and cognitive determinants of chatbot humanness on learning outcomes and motivation

Dependent variables		**Condition**	**Mean**	**Standard deviation**	**F-value** $F(1, 51)$	**p-value**	**Effect size**
Learning outcome		CH	5.20	2.08	1.35	0.25	0.03
		CN	5.69	1.60			
		AH	3.07	2.02			
		AN	4.77	1.92			
Learning motivation	1	CH	5.69	1.03	1.14	0.29	0.02
		CN	6.01	0.83			
		AH	5.47	0.90			
		AN	5.25	0.97			
	2	CH	2.63	1.41	5.16	0.03*	0.09

		CN	2.22	1.12			
		AH	2.03	1.07			
		AN	3.17	1.44			
	3	CH	4.25	0.88	0.05	0.82	1.05e-03
		CN	4.46	0.70			
		AH	4.63	0.90			
		AN	4.73	0.87			
	4	CH	4.83	1.24	4.94	0.03*	0.09
		CN	5.51	0.66			
		AH	5.36	0.95			
		AN	4.85	0.99			
	5	CH	5.53	1.15	1.08	0.30	0.02
		CN	6.00	0.76			
		AH	5.84	0.83			
		AN	5.77	1.00			

$p^* < 0.05$

5.4.2. Qualitative Data Analysis

Despite the lack of explicit evidence that can directly explain the quantitative results, there were several comments that might be able to explain why tension-pressure and perceived competence dimensions of learning motivation were impacted by the interaction between cognitive engagement mode of learning and cognitive determinants of chatbot humanness. Concerning the tension-pressure dimension of learning motivation, three constructive & humanized chatbot learners (CH_8, 9, 11) mentioned that they felt pressure or tension while answering their opinions, and one active & non-humanized chatbot learner (AN_4), while two constructive & non-humanized chatbot learners (CN_1,2) reported that they did not feel like talking with a human, thus not feeling pressure. In contrast, one active & non-humanized chatbot learners (AN_14) reported that she kept feeling pressure while selecting the correct answer out of several options.

> *"I hardly felt any pressure, but I felt a bit anxious and burdened because there were many questions asking about my own thoughts." (CH_9)*
>
> *"When I had to answer questions that didn't have a clearcut answer but asked for my opinions, I felt burdened because I had to think a lot." (CH_8)*
>
> *"I didn't feel like there was a person waiting in front of me; instead, the chatbot just provided the problems and hints, allowing me to think independently and without any pressure, which I found to be beneficial." (CN_1)*
>
> *"I might have felt tension a bit, but no pressure at all. I just enjoyed learning because it felt like playing a quiz game." (CN_2)*
>
> *"When addressing ethical issues, I think it is essential to deliberate and contemplate. It would have been better if I had a counterpart with whom I can share such considerations and thoughts. When I have to solve problems that can benefit from discussion, I want to discuss with a chatbot." (CH_1)*
>
> *"The process of solving problems with the chatbot wasn't particularly enjoyable for me. It felt more like striving to get the correct answers rather than enjoying the journey itself." (AN_14)*

Regarding the perceived competence dimension of learning motivation, we could not find comments that can directly explain why learners perceived themselves as being competent or not. The followings are example comments that might be able to explain the unexpected interaction effect on the perceived competence, which is discussed in subsection 6.3.

> *"I preferred solving problems asking my thoughts to writing predetermined answer, because it was more interesting to express my opinion or make inferences." (CN_8, 13)*
>
> *"After watching a video or doing a simulation, in which I didn't have to put much effort understanding, I had to think more by myself when solving problems." (CH_3, 5 & CN_9)*
>
> *"I ended up thinking further what I learned while applying it to other cases." (CH_6, 13)*

Chapter 6. Discussion

According to quantitative analysis results, learning activities designed to induce a constructive level of cognitive engagement led learners to higher learning outcomes but not to higher learning motivation. Chatbots designed to be perceived as machinelike led learners to higher learning outcomes but not to higher learning motivation than humanized chatbots did. Such results were supported by quantitative analysis results, while some qualitative analysis results were not aligned with quantitative analysis results.

6.1. H1: Main Effect of Cognitive Engagement Mode of Learning

As hypothesized in H1a, constructive chatbot learners achieved a higher level of learning outcomes than those using active chatbot learners did. In addition to the statistically significant results, we could observe that some active chatbot learners made mistakes in the last question, which was not observed in the constructive chatbot learners. Considering that the majority of active chatbot learners evaluated the difficulty level of problem-solving as no greater than medium, the active chatbot learners' engagement in learning might have decreased as they proceeded to select and give short answers, while constructive chatbot learners had to write longer sentences.

Thus, such results indicate that the learning outcome predictions based on the ICAP framework can also be applied to problem-solving-focused learning using chatbots. Meanwhile, according to the ICAP framework, exceptional cases were reported where learning outcomes in constructive learning were not significantly different from those in active learning, especially when learners lacked knowledge schemas. The context of this study was such that the learners had little prior knowledge about artificial intelligence, and none had experience using text-based chatbots for educational purposes. Therefore, the results predicted in H1a might not have emerged. However, in a context similar to this study (a basic level AI education for non-expert adults with no background knowledge), it seems more desirable to aim for inducing constructive learning rather than active learning to improve learning outcomes.

Contrary to the hypothesis, no significant difference was reported in learning motivation between constructive chatbot learners and active chatbot learners. Through the post-interview, we were able to elucidate the cause of this non-significant difference partially. Five constructive chatbot learners reported that the problems were too difficult. While constructive learning had a positive impact on learning outcomes, it might not have had such a positive effect on learning motivation, particularly interest-enjoyment. However, four constructive chatbot learners (CN_8, 13 & CH_3, 13) commented that constructive questions were intriguing and interesting, as these questions provoked them to reorganize what they learned. Considering these points, learning motivation appears less influenced by adjusting cognitive engagement levels than learning outcomes. For enhancing learning motivation, it may be worth considering the regulation of emotional engagement level, another type of learner engagement.

In sum, we concluded that learning activities demanding high cognitive engagement alone could yield positive effects on learning outcomes, but they did not make a noticeable difference in learning motivation. If the primary learning goal is set as the achievement of a high learning outcome, it would be desirable to design a chatbot-based learning curriculum where learning activities are more assigned for a constructive mode of learning rather than an active mode of learning. On the other hand, if considering learning motivation as the main goal, it may be necessary to consider factors beyond the manipulation of cognitive engagement level.

6.2. H2: Main Effect of Cognitive Determinants of Chatbot Humanness

Contrary to what was anticipated in H2a, the learning outcomes of non-humanized chatbot learners were higher than those of humanlike chatbot learners. One possible explanation for this counter-hypothesis result could be the enhanced advantages of guided learning. Regardless of using cognitive determinants of chatbot humanness, a significant number of learners commented that the chatbot led and guide the entire learning process, which was helpful for learning knowledge of a new field. Given that the strong leadership of the chatbot and the established step-by-step curriculum were uniformly applied to all four experimental groups, they would not have directly influenced the learning outcomes. However, such benefits of chatbot-guided learning could potentially have been further amplified by the fact that the non-humanized chatbot used a more formal speech style and did not engage in small talk.

Moreover, although it does not explain the results that contradict H2a, there was a sentiment among non-humanized chatbot learners that they perceived their interaction with the chatbot as humanlike, and they found this beneficial to their learning. Interestingly, this was similar to reports from humanized chatbot learners.

For example, they mentioned that interacting with the chatbot felt like interacting with a human, or they experienced a sense of social presence as if they were with another person. This might suggest that the dispositional determinants of perceived humanness led some non-humanized chatbot learners to perceive their interaction with the chatbot as more humanlike and natural. Such positive perception of interaction with the chatbot may also have had a positive impact on learning outcomes.

Contrary to what was anticipated in H2b, the cognitive determinants of chatbot humanness did not have a significant main effect on learning motivation. Two possible reasons were found. Firstly, both experimental groups (i.e., humanized and non-humanized chatbots) might have demonstrated high motivation, considering that the learners had little prior knowledge and learned basic information about artificial intelligence for the first time. In fact, looking at the statistical report in Table 9, both groups scored above the median of 4 on the Likert scale for all dimensions of learning motivation except the second one. Additionally, given that most adults typically had experience in solving problems on paper but no prior experience in using chatbots for educational purposes such as problem-solving, this novel approach could have led to high learning motivation, which was observed uniformly across the experimental groups.

> *"Indeed, it was somewhat more interesting to solve problems through the chatbot than just writing the answers on paper."(CN_6)*
>
> *"Compared to just reading a book, learning with the chatbot was much more enjoyable." (CH_5)*

6.3. H3: Interaction Effect

Both humanized and non-humanized chatbots had a main effect, but there was no interaction effect on learning outcomes. However, for tension-pressure (the second dimension of learning motivation), there was an interaction effect as predicted in H3b. In other words, for constructive chatbot learners, when the chatbot was humanized, tension-pressure was higher, and perceived competence was lower than when it was not. For active chatbot learners, conversely, when the chatbot was not humanized, tension-pressure was lower, while perceived competence was higher. We analyzed that these two interaction effects originated from the fact that learners in constructive learning became more engaged in learning, thus being more greatly influenced by the main effect of the cognitive determinants of chatbot humanness (either a positive or negative effect), than learners in active learning. Thus, the effect of cognitive engagement level on learning motivation should be qualified depending on whether the the cognitive determinants of chatbot humanness is employed or not. Based on the interview responses from the participants, which were described in subsection 5.5.2., we discussed the detailed causes as follows.

Regarding the tension-pressure dimension, participants reported that activities involving expressing their own opinions and thoughts in constructive learning brought about tension-pressure. However, despite the fact that both constructive & humanized and constructive & non-humanized chatbot learners engaged in the constructive learning activities in the same way, whether such activities led to tension-pressure seemed to differ depending on the presence of the cognitive determinants of chatbot humanness. Humanized chatbot learners may have felt that the interaction with the chatbot was similar to an interaction with a human, and, therefore, could have felt more pressure when justifying their argument. On the other hand, non-humanized chatbot learners, who considered the chatbot as non-human, may have felt a lower degree of tension or pressure during the same constructive learning activities.

Concerning the perceived competence dimension of learning motivation, constructive & humanized learners perceived a lower level of competence than constructive & non-humanized chatbot learners. One possible explanation can be a difference in expectations toward constructive chatbots between humanized and non-humanized chatbot learners. Both humanized and non-humanized chatbot learners reported that they benefited from constructive learning, where they could engage in deep thinking. According to Jiang et al. (2023), when chatbots used humanlike cues, users perceived such chatbots as more competent at task-solving than chatbots not using such cues. Such perception of competence toward chatbots led to increased trust on the chatbots. Based on this finding, constructive & humanized chatbot learners might have expected a higher level of competence in chatbots, thus expecting support in their problem-solving. However, constructive chatbots were designed to support learners in solving problems independently to show deep cognitive engagement. Such a lack of active role of chatbot during the problem-solving process might have led to low perceived competence.

Chapter 7. Conclusion

This study aimed to examine the impact of a learning experience-related factor (i.e., cognitive engagement in learning) and a user experience-related factor (i.e., cognitive determinants of chatbot humanness) on learning outcomes and motivation. Four versions of chatbots (i.e., constructive & humanized, constructive & non-humanized, active & humanized, active & non-humanized chatbots) were designed, based on the ICAP framework and the CMD framework. Constructive chatbots lead learners to (1) explain or summarize concepts in the video/simulation in one's own word; (2) justify my opinion based on what s/he learned; (3) infer new information from what was explicitly taught; (4) generate predictions on a new case that was not addressed during learning. In contrast, active chatbots require learners to fill in the blank of the full sentences or select a correct answer among the various options, without need to generate a new output or infer new information. Humanized chatbots were designed by using six cognitive determinants of perceived humanness: (1) provide visual representation of humanlike character, (2) identify a chatbot's role and name as humanlike, (3) use social dialogues, (4) use informal and casual verbal style, (5) use emojis, and (6) use stickers that have facial expressions.

Based on mixed-method data analysis, we showed that there were the main effects of cognitive engagement mode of learning and cognitive determinants of chatbot humanness on learning outcomes, while there were no main effects on learning motivation. The interaction effect was detected only on learning motivation (tension-pressure and perceived competence).

Our study has three limitations, which lead to suggestion of future study. First, response data might not be sufficient to reflect the entire aspects of cognitive engagement in learning. It would be interesting to measure cognitive engagement in learning by employing diverse approach, such as self-report data and multi-modal data collection. Second, our study did not confirm which cognitive determinant gave a more significant impact on learning outcomes/ motivation than other determinants. The post-interview data showed limited findings on such impact of each determinant but did not prove such findings based on a quantitative analysis approach. A future study can be designed to quantitatively analyze which cognitive determinants of chatbot humanness are critical in learning effects. Lastly, since our user experiment was conducted for one-time, we could not confirm whether our study's results will be effective in a longer term and a real educational setting. Such a long-term effect can be explored through another vein of future study, such as longitudinal study and an in-situ user experiment.

Despite the limitations, our study has four-fold contributions: (1) We designed our educational chatbot based on the concepts of learning experience design to encompass both the instructional design element and the user experience element; (2) Despite the learners lacked background knowledge, constructive chatbot were more effective in improving learning outcomes than active chatbots; (3) Different from our prediction, which was based on previous literature, non-humanized chatbots were more effective in learning outcomes than humanized chatbots; (4) When using constructive chatbots, cognitive determinants of chatbot humanness gave different impacts on two types of learning motivation: positive impact on tension-pressure and negative impact on perceived competence.

References

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Computers in Human Behavior, 85, 183–189.

Argyle, M. (2013). Bodily Communication. Routledge.

Ayedoun, E., Hayashi, Y., & Seta, K. (2015). A Conversational Agent to Encourage Willingness to Communicate in the Context of English as a Foreign Language. Procedia Computer Science, 60, 1433–1442.

Bai, Q., Dan, Q., Mu, Z., & Yang, M. (2019). A Systematic Review of Emoji: Current Research and Future Perspectives. Frontiers in Psychology, 10, 2221.

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. Computers in Human Behavior, 25(2), 450–457.

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A Bot and a Smile: Interpersonal Impressions of Chatbots and Humans Using Emoji in Computer-mediated Communication. Communication Studies, 71(3), 409–427.

Benbasat, I., Dimoka, A., Pavlou, P. A., & Qiu, L. (2010). Incorporating Social Presence in the Design of the Anthropomorphic Interface of Recommendation Agents: Insights from an fMRI Study. https://aisel.aisnet.org/icis2010_submissions/228/

Bickmore, T., & Cassell, J. (2001). Relational agents: a model and implementation of building user trust. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 396–403.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput.-Hum. Interact., 12(2), 293–327.

Billett, S. (2012). Guided Learning. In N. M. Seel (Ed.), Encyclopedia of the Sciences of Learning (pp. 1403–1406). Springer US.

Bonwell, C. C., & Eison, J. A. (1991). Active learning: Creating excitement in the classroom. 1991 ASHE-ERIC higher education reports. ERIC Clearinghouse on Higher Education, The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183. https://files.eric.ed.gov/fulltext/ED336049.pdf

Brody, N., & Caldwell, L. (2019). Cues filtered in, cues filtered out, cues cute, and cues grotesque: Teaching mediated communication with emoji Pictionary. *Communication Teacher*, 33(2), 127–131.

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, 90, 315–330.

Cha, Y., Kim, J., Park, S., Yi, M. Y., & Lee, U. (2018). Complex and Ambiguous: Understanding Sticker Misinterpretations in Instant Messaging. *Proc. ACM Hum.-Comput. Interact., 2(CSCW)*, 1–22.

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.

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. Educational Psychologist, 49(4), 219–243.

Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J., & Popescu, V. (2017). Hand Gesture and Mathematics Learning: Lessons From an Avatar. Cognitive Science, 41(2), 518–535.

Corbin, J., & Strauss, A. (2015). Basics of Qualitative Research. SAGE.

Cowell, A. J., & Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. International Journal of Human-Computer Studies, 62(2), 281–306.

Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. Computers & Education, 190, 104607.

Deveci Topal, A., Dilek Eren, C., & Kolburan Geçer, A. (2021). Chatbot application in a 5th grade science course. Education and Information Technologies, 26(5), 6241–6265.

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. Journal of the Association for Information Systems, 23(1), 96–138.

Driscoll, M., & Burner, K. (2019). *Psychology of Learning for Instruction*. Pearson.

Dunlap, J. C., Bose, D., Lowenthal, P. R., York, C. S., Atkinson, M., & Murtagh, J. (2016). What Sunshine Is to Flowers. In Emotions, Technology, Design, and Learning (pp. 163–182).

Elder, A. M. (2018). What Words Can&t Say. Journal of Information Communication and Ethics in Society, 16(1), 2–15.

Eom, S. B., Wen, H. J., & Ashill, N. (2006). The determinants of students& perceived learning outcomes and satisfaction in university online education: An empirical investigation. Decision Sciences Journal of Innovative Education, 4(2), 215–235.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886.

Fane, J., MacDougall, C., Jovanovic, J., Redmond, G., & Gibbs, L. (2018). Exploring the use of emoji as a visual research method for eliciting young children&s voices in childhood research. Early Child Development and Care, 188(3), 359–374.

Farah, J. C., Sharma, V., Ingram, S., & Gillet, D. (2021). Conveying the Perception of Humor Arising from Ambiguous Grammatical Constructs in Human-Chatbot Interaction. Proceedings of the 9th International Conference on Human-Agent Interaction, 257–262.

Farah, J. C., Spaenlehauer, B., Sharma, V., Rodríguez-Triana, M. J., Ingram, S., & Gillet, D. (2022). Impersonating Chatbots in a Code Review Exercise to Teach Software Engineering Best Practices. 2022 IEEE Global Engineering Education Conference (EDUCON), 1634–1642.

Fidan, M., & Gencel, N. (2022). Supporting the Instructional Videos With Chatbot and Peer Feedback Mechanisms in Online Learning: The Effects on Learning Performance and Intrinsic Motivation. Journal of Educational Computing Research, 60(7), 1716–1741.

Floor, N. (2023). This Is Learning Experience Design: What It Is, How It Works, and Why It Matters. Pearson Education.

Franke, T., Attig, C., & Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. International Journal of Human–Computer Interaction, 35(6), 456–467.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44068/1/10_03_emerging.pdf

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. In Computers in Human Behavior (Vol. 75, pp. 461–468). https://doi.org/ 10.1016/j.chb.2017.05.045

Gnewuch, U., Morana, S., Adam, M., & Maedche, A. (2018). Faster is not always better: understanding the effect of dynamic response delays in human-chatbot interaction. https://scholar.archive.org/work/664dlhdlrbdvffn4jb4zir4toa/access/wayback/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1112&context=ecis2018_rp

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior, 97, 304–316.

Gong, L., & Nass, C. (2007). When a Talking-Face Computer Agent is Half-Human and Half-Humanoid: Human Identity and Consistency Preference. Human Communication Research, 33(2), 163–193.

Haristiani, N. (2019). Artificial Intelligence (AI) Chatbot as Language Learning Medium: An inquiry. Journal of Physics. Conference Series, 1387(1), 012020.

Harrington, C. N., & Egede, L. (2023, April 19). Trust, comfort and relatability: Understanding black older adults& perceptions of chatbot design for health information seeking. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI &23: CHI Conference on Human Factors in Computing Systems, Hamburg Germany. https://doi.org/ 10.1145/3544548.3580719

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., Luo, J., & Akkiraju, R. (2018). Touch Your Heart: A Tone-aware Chatbot for Customer Care on Social Media. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Article Paper 415.

Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. Interactive Learning Environments, 1–14.

Jain, M., Kumar, P., Bhansali, I., Liao, Q. V., Truong, K., & Patel, S. (2018). FarmChat: A Conversational Agent to Answer Farmer Queries. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2(4), 1–22.

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. Proceedings of the 2018 Designing Interactive Systems Conference, 895–906.

Janssen, J. H., IJsselsteijn, W. A., & Westerink, J. H. D. M. (2014). How affective technologies can influence intimate interactions and improve social connectedness. International Journal of Human-Computer Studies, 72(1), 33–43.

Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: young learners& experiences and perspectives. Computer Assisted Language Learning, 1–26.

Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. Knowledge-Based Systems, 22(4), 249–255.

Jia, J., Chen, Y., Ding, Z., & Ruan, M. (2012). Effects of a vocabulary acquisition and assessment system on students& performance in a blended learning class for English subject. Computers & Education, 58(1), 63–76.

Jiang, H., Cheng, Y., Yang, J., & Gao, S. (2022). AI-powered chatbot communication with customers: Dialogic interactions, satisfaction, engagement, and customer behavior. Computers in Human Behavior, 134, 107329.

Jiang, Y., Yang, X., & Zheng, T. (2023). Make chatbots more adaptive: Dual pathways linking human-like cues and tailored response to trust in interactions with chatbots. Computers in Human Behavior, 138, 107485.

Jin, L., Wen, Z., & Gough, N. (2010). Social virtual worlds for technology-enhanced learning on an augmented learning platform. Learning, Media and Technology, 35(2), 139–153.

Kim, N.-Y., Cha, Y., & Kim, H.-S. (2019). Future English learning: Chatbots and artificial intelligence. Multimedia-Assisted Language Learning, 22(3), 32–53.

Kim, S., Eun, J., Oh, C., Suh, B., & Lee, J. (2020). Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13.

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? Computers in Human Behavior, 28(1), 241–250.

Konya-Baumbach, E., Biller, M., & von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. Computers in Human Behavior, 139, 107513.

Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. Computers & Education, 191, 104646.

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.

Li, W., Wang, F., Mayer, R. E., & Liu, T. (2022). Animated pedagogical agents enhance learning outcomes and brain activity during learning. *Journal of Computer Assisted Learning*, 38(3), 621–637.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science, 4. https://www.researchgate.net/profile/Karl-Macdorman/publication/241217609_Subjective_Ratings_of_Robot_Video_Clips_for_Human_Likeness_Familiarity_and_Eeriness_An_Exploration_of_the_Uncanny_Valley/links/00b7d5320f9213af36000000/Subjective-Ratings-of-Robot-Video-Clips-for-Human-Likeness-Familiarity-and-Eeriness-An-Exploration-of-the-Uncanny-Valley.pdf

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Research Quarterly for Exercise and Sport, 60(1), 48–58.

Milcent, A.-S., Kadri, A., & Richir, S. (2022). Using Facial Expressiveness of a Virtual Agent to Induce Empathy in Users. International Journal of Human–Computer Interaction, 38(3), 240–252.

Moffitt, R. L., Padgett, C., & Grieve, R. (2020). Accessibility and emotionality of online assessment feedback: Using emoticons to enhance student perceptions of marker competence and warmth. Computers & Education, 143, 103654.

Nguyen, H. (2022). Exploring group discussion with conversational agents using epistemic network analysis. In Communications in Computer and Information Science (pp. 378–394). Springer International Publishing.

Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking. Computers & Education, 192, 104661.

Nguyen, H. D., Pham, V. T., Tran, D. A., & Le, T. T. (2019). Intelligent tutoring chatbot for solving mathematical problems in High-school. *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, 1–6.

Parrish, P. E. (2009). Aesthetic principles for instructional design. *Educational Technology Research and Development: ETR & D*, 57(4), 511–528.

Piaget, J., & Cook, M. (1952*). The origins of intelligence in children* (Vol. 8, No. 5, pp. 18-1952). New York: International Universities Press.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2168-9830.2004.tb00809.x

Qiu, L., & Benbasat, I. (2010). A study of demographic embodiments of product recommendation agents in electronic commerce. *International Journal of Human-Computer Studies*, 68(10), 669–688.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents. In arXiv [cs.CY]. arXiv. http://arxiv.org/abs/1704.04579

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, 102630.

Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior*, 126, 107034.

Roberta, D. C., Silva, S. C., & Romana, A. F. (2020). Millennials& attitude toward chatbots: an experimental study in a social relationship perspective. International *Journal of Retail & Distribution Management*, 48(11), 1213–1233.

Rocca, K. A., & McCroskey, J. C. (1999). The interrelationship of student ratings of instructors& immediacy, verbal aggressiveness, homophily, and interpersonal attraction. *Communication Education*, 48(4), 308–316.

Roy, R., & Naidoo, V. (2021). Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 126, 23-34.

Rubin, V. L., Chen, Y., & Marie, T. L. (2010). Artificially intelligent conversational agents in libraries. Library Hi Tech, 28(4), 496–522.

Sah, Y. J., & Peng, W. (2015). Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. Computers in Human Behavior, 45, 392–401.

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with Humanlike Conversational Agents: Designing for Anthropomorphism. Journal of the Association for Information Systems, 22(4), 8.

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots:

Anthropomorphism and adoption. Journal of Business Research, 115, 14–24.

Skjuve, M., & Brandtzæg, P. B. (2018). Chatbots as a new user interface for providing health information to young people. Nordicom.

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. Computers & Education, 151, 103862.

Sundar, S. S. (2008). Self as source: Agency and customization in interactive media. In *Mediated interpersonal communication* (pp. 72-88). Routledge.

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers & Education, 50(4), 1183–1202.

Tang, Y., & Hew, K. F. (2019). Emoticon, Emoji, and Sticker Use in Computer-Mediated Communication: A Review of Theories and Research Findings. International Journal of Communication Systems, 13(0), 27.

Vareberg, K. R., Vogt, O., & Berndt, M. (2022). Putting your best face forward: How instructor emoji use influences students& impressions of credibility, immediacy, and liking. Education and Information Technologies, 1–18.

Wadsworth, B. J. (1996). Piaget&s theory of cognitive and affective development: Foundations of constructivism, 5th ed. 5, 195.

Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Article Article 683.

Wambsganss, T., Söllner, M., & Leimeister, J. M. (2020). Design and Evaluation of an Adaptive Dialog-Based Tutoring System for Argumentation Skills. https://papers.ssrn.com/abstract=3911888

Winkler, R., Hobert, S., Fischer, T., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Engaging Learners in Online Video Lectures with Dynamically Scaffolding Conversational Agents. https://papers.ssrn.com/abstract=3915629

Winkler, R., Söllner, M., & Leimeister, J. M. (2021). Enhancing problem-solving skills with smart personal assistant technology. Computers & Education, 165, 104148.

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2011). Vancouver, British Columbia (May 7-12, 2011). New York: ACM Press, pp. 143-146. https://depts.washington.edu/acelab/proj/art/. DOI: 10.1145/1978942.1978963.

Woo, H. L. (2009). Designing multimedia learning environments using animated pedagogical agents: factors and issues. Journal of Computer Assisted Learning, 25(3), 203–218.

Wuenderlich, N. V. and Paluch, S., "A Nice and Friendly Chat with a Bot: User Perceptions of AI-Based Service Agents" (2017). ICIS 2017 Proceedings. 11.

Xu, W., Ma, J., Yao, J., Lin, W., Zhang, C., Xia, X., Zhuang, N., Weng, S., Xie, X., Feng, S., Ying, F., Hansen, P., & Yao, C. (2023). MathKingdom: Teaching Children Mathematical Language Through Speaking at Home via a Voice-Guided Game. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Article Article 93.

Yin, J., Goh, T.-T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students& learning motivation and performance. ACM Journal on Educational Resources in Computing, 59(1), 154–177. Zhou, R., Hentschel, J., & Kumar, N. (2017). Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 748–759.
Zhou, R., Hentschel, J., & Kumar, N. (2017). Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 748–759.

Appendix

Appendix #1. Descriptive statistics of age, pre-test score, Affinity for Technology Interaction (ATI) score, and gender ratio across the Four Experimental Conditions

		Age	Pre-test score	ATI score	Gender ratio
Active & Humanized	*m*	31.93	1.93	3.59	Female (n = 11)
	sd	9.51	1.33	0.77	Male (n = 3)
Active & Non-humanized	*m*	28.85	2.15	3.65	Female (n = 10)
	sd	7.68	1.86	0.69	Male (n = 3)
Constructive & Humanized	*m*	32.93	2.47	3.41	Female (n = 12)
	sd	8.40	1.13	0.52	Male (n = 3)
Constructive. & Non-humanized	*m*	29.30	2.08	3.62	Female (n = 10)
	sd	5.64	0.86	0.63	Male (n = 3)

Appendix #2. Problem-solving activities (Korean)

No.	Constructive chatbot	Active chatbot
0-1	인공지능이 무엇인지 "인공"과 "지능"이라는 단어를 사용하여 설명하시오.	인공지능이 무엇인지 설명하는 다음 문장의 빈칸에 들어갈 알맞은 단어를 골라보시오. <문장> ______(을)를 인공적으로 흉내내는 소프트웨어, 프로그램, 장치 모두를 인공지능이라고 한다.
1-1	기계학습이 무엇인지 "기계"와 "학습"이라는 단어를 사용하여 설명하시오.	기계학습이 무엇인지 설명하는 다음 문장의 빈칸에 들어갈 알맞은 단어 조합을 골라보시오. <문장> 기계학습이란 [데이터, 사람] (을)를 사용하여 기계가 [스스로, 의존적으로] 학습하게 하는 방법이다.
1-2	기존의 컴퓨터 프로그램과 기계학습의 차이를 설명하시오.	기존의 컴퓨터 프로그램과 기계학습의 차이를 설명하는 다음 문장의 빈칸에 공통적으로 들어갈 단어를 쓰시오. <문장> 기존의 컴퓨터 프로그램에서는 사람이 직접 컴퓨터에 필요한 ____(을)를 입력한다. 반면 기계학습의 경우는 사람이 아닌 컴퓨터가 직접 수많은 데이터를 분석해서 알맞은 ____(을)를 스스로 찾아낸다.
1-3	기계학습은 다양한 데이터에서 패턴을 인식하는 법을 배웁니다. 예를 들어, 어떤 기계학습이 자동차	기계학습은 다양한 데이터에서 패턴을 인식하는 법을 배웁니다. 예를 들어, 어떤 기계학습이 내가 좋아하는 동영상과 좋아하지 않는 동영상의 차이를 학습했

	사진들과 자전거 사진들의 차이를 배웠다고 가정합시다. 이러한 학습을 통해 이 기계학습이 궁극적으로 수행하고자 하는 것은 무엇인지 설명하시오.	다고 가정합시다. 이러한 학습을 통해 이 기계학습이 궁극적으로 수행하고자 하는 것은 무엇인지 다음 문장의 빈칸에 들어갈 알맞은 단어 조합을 고르시오. <문장> 이 기계학습은 어떤 [새로운/기존의] 동영상이 주어졌을 때, 내가 좋아할 영상인지 아닌지 [예측/상상]하는 것을 궁극적인 목표로 한다.
2-1	기계학습을 학습시키기 위해서 사람이 해야하는 작업 과정이 있습니다. 그 과정을 순차적으로 설명하시오.	기계학습을 학습시키기 위해서 사람이 해야하는 작업 과정이 있습니다. 이 과정을 알맞게 설명한 문장이 몇 번인지 고르시오. (1번) 주어진 데이터로부터 라벨을 제거하고, 라벨로 인공지능을 학습시킨다. (2번) 주어진 데이터에 라벨을 달아주고, 라벨로 인공지능을 학습시킨다. (3번) 주어진 데이터에 라벨을 제거하고, 그 데이터로 인공지능을 학습시킨다. (4번) 주어진 데이터에 라벨을 달아주고, 그 데이터로 인공지능을 학습시킨다.
2-2	만약 인간이 데이터에 잘못된 라벨을 달아 인공지능에게 제공하면 어떤 일이 발생할지, 왜 그런지 설명하시오.	만약 인간이 실수로 데이터에 잘못된 라벨을 달아 인공지능에게 제공하면 어떤 문제이 발생할지, 왜 그런지 설명하는 다음 문장에 들어갈 알맞은 단어를 적어주세요. <문장> 잘못된 라벨이 달린 데이터로 학습한 인공지능은 잘못된 ____(을)를 반복적으로 할 것이다.
2-3	인공지능이 새로운 데이터	인공지능이 새로운 데이터를 보다 정확

	를 보다 정확하게 분류하기 위해서 필요한 조건이 2가지 있습니다. 2가지 조건이 무엇인지 이번 학습단계에서 여러분이 인공지능을 학습시킬 때의 경험을 바탕으로 설명해보시오. 힌트: 데이터, 양, 라벨(혹은 이름표), 학습	하게 분류하기 위해서 필요한 조건이 2가지 있습니다. 이번 학습단계에서 여러분이 인공지능을 학습시킬 때의 경험을 바탕으로 2가지 조건이 무엇인지 설명하기 위해 다음 문장의 빈칸에 알맞은 단어 조합을 고르시오. <문장> (1) 인공지능이 [많은 양의/선별된 소수의] 데이터를 학습하는 것 (2) 인공지능이 [간단한/올바른] 라벨이 달린 데이터를 학습하는 것
3-1	우리가 방금 학습시킨 이 인공지능을 바닷 속에 무엇이 있어야 할지 결정할 때 사용할 수 있는지에 대한 본인의 주장과 근거를 설명하시오.	우리가 방금 학습시킨 이 인공지능을 바닷 속에 무엇이 있어야 할지 결정할 때 사용할 수 있는지 설명하는 다음 문장의 빈칸에 각각 들어갈 두 단어를 답하시오. <문장> 사용하기에 적절하지 않다. 왜냐하면 인공지능(AI)이 ______(이)가 아닌 ________(을)를 쓰레기로 분류하여 의도하지 않게 이들을 죽일 수 있기 때문이다.
3-2	우리가 방금 학습시킨 이 인공지능을 사용하여 바닷 속에 무엇이 있어야 할지 결정한다면, 인공지능이 물고기가 아닌 해양 생물들을 쓰레기로 분류하여 의도하지 않게 이들을 죽일 수 있습니다. 이러한 문제를 어떻게 해결할 수 있는	우리가 방금 학습시킨 이 인공지능을 사용하여 바닷 속에 무엇이 있어야 할지 결정한다면, 인공지능이 물고기가 아닌 해양 생물들을 쓰레기로 분류하여 의도하지 않게 이들을 죽일 수 있습니다. 이러한 문제를 어떻게 해결할 수 있는 방법으로 가장 적절한 것을 보기 중에서 고르시오. <보기>

	방법을 한 가지 설명하시오.	(1번) 인공지능에게 더 많은 데이터를 제공해서 정확한 판단을 내리게 한다. (2번) 인공지능이 학습할 데이터의 양을 줄여서 학습의 부담을 줄여준다. (3번) 인공지능이 물고기, 물고기가 아닌 해양생물, 쓰레기를 모두 구분하도록 학습시킨다. (4번) 인공지능만으로는 분류가 어려우므로 사람이 함께 물고기와 물고기가 아닌 해양생물을 구분해준다.
4-1	이번 단계의 인공지능이 앞서 학습시킨 인공지능과 어떻게 다르게 학습되었는지 설명하시오.	다음 보기 중 이번 단계의 인공지능이 앞서 학습시킨 인공지능과 어떻게 다르게 학습되었는지를 가장 적절하게 설명하는 것을 고르시오. <보기> (1번) 이번 단계의 인공지능은 보다 많은 양의 데이터를 학습했다. (2번) 이번 단계의 인공지능은 보다 적은 양이지만 고화질의 데이터를 학습했다. (3번) 이번 단계의 인공지능은 물고기와 물고기가 아닌 해양생물 모두가 쓰레기가 아닌 것으로 분류하도록 학습했다. (4번) 이번 단계의 인공지능은 물고기와 물고기가 아닌 해양생물 모두가 쓰레기로 분류하도록 학습했다.
4-2	지금까지 우리가 인공지능을 학습시킨 방식은 지도학습, 비지도학습, 강화학습 중에 무엇에 해당하는지 여러분의 주장과 근거를 설명하시오.	지금까지 우리가 인공지능을 학습시킨 방식은 지도학습, 비지도학습, 강화학습 중에 무엇에 해당하는지 설명하는 다음 문장의 빈칸에 각각 들어갈 알맞은 말을 답하시오. <문장>

		인공지능이 물고기인지, 물고기가 아닌 해양생물인지, 쓰레기인지 ____(을)를 달아준 데이터를 학습했기 때문에 ________에 해당합니다.
5-1	지도학습이란 무엇인지 "지도"와 "학습"이라는 단어를 모두 사용하여 설명하시오.	지도학습이란 무엇인지 설명하는 다음 문장의 빈칸에 들어갈 알맞은 단어의 조합을 고르시오. <문장> 지도학습이란 기계학습의 종류 중 하나이며, [기계/인간] (이)가 인공지능에게 데이터의 [정답/용량]을 지도하여 학습시키는 방법이다. 1번) 기계, 정답 2번) 기계, 용량 3번) 인간, 정답 4번) 인간, 용량
5-2	이메일 서비스에서 사용되는 기계학습은 사용자에게 어떤 메일이 스팸메일인지 아닌지 구분해줄 수 있습니다. 이와 비슷한 원리로, 유튜브에서 활용되는 기계학습이 사용자에게 다음에 시청할 동영상을 추천해주는 과정을 다음 단어들을 활용하여 설명하시오. <단어> 레이블(정답), 사람, 기계학습, 학습	이메일 서비스에서 사용되는 기계학습은 사용자에게 어떤 메일이 스팸메일인지 아닌지 구분해줄 수 있습니다. 기계학습이 스팸메일을 분류하는 과정을 설명하도록 다음 문장들을 올바른 순서로 나열하시오. <문장> (ㄱ) 사람이 어떤 메일이 스팸이고 아닌지 레이블(정답)을 데이터에 달아준다. (ㄴ) 학습을 바탕으로 새로운 메일이 스팸인지 아닌지 판단한다. (ㄷ) 기계학습이 어떤 메일이 스팸메일 레이블에 포함되는지 학습한다.
6-1	이번 단계에서 인공지능은 어떻게 프로그래밍 되었고,	이번 단계에서 인공지능은 어떻게 프로그래밍 되었고, 이는 앞서 프로그래밍

	이는 앞서 프로그래밍 한 인공지능과 어떻게 다른지 설명하시오.	한 인공지능과 어떻게 다른지 설명하는 다음 문장의 빈칸 2개에 각각 들어갈 두 단어를 답하시오. <문장> 학습데이터를 활용해 AI가 ____물고기를 인식하도록 프로그래밍했습니다. 물고기와 물고기가 아닌 것들을 구분하는 것에서 더 나아가 물고기를 ___에 따라 분류할 수 있게 학습하였습니다.
7-1	비지도학습이 무엇인지 "지도"와 "학습"이라는 단어를 모두 사용하여 설명하시오.	비지도학습이란 무엇인지 설명하는 다음 문장의 빈칸에 들어갈 알맞은 단어의 조합을 고르시오. <문장> 비지도학습은 기계학습 중 하나이며, 사람이 [결과/정답] (을)를 알려주지 않아도 인공지능이 여러 [레이블/데이터] 속에서 관계나 패턴을 찾아 스스로 학습하는 방법이다.
7-2	수만명의 학습자가 사용하는 온라인 학습공간에서 학습자의 학습 성향을 분석해야 하는 상황을 생각해봅시다. 이런 상황에서 지도학습과 비지도학습 중 어떤 방법이 더 적절할지 주장과 근거를 설명해주세요.	많은 양의 뉴스를 비슷한 뉴스끼리 그룹으로 묶어주어야 하는 상황을 생각해봅시다. 이런 상황에서 지도학습과 비지도학습 중 어떤 방법이 더 적절할지 설명하는 다음 문장의 빈칸에 들어갈 알맞은 단어 조합을 골라봅시다. <문장> [지도학습/비지도 학습]이 더 적절하다. 왜냐하면 이 방법이 [흥미로운 특징/분명한 정답] 이 없는 많은 데이터를 학습하는데 유리하기 때문이다.
7-3	지도학습과 비지도학습에서 사람이 하는 역할의 차	지도학습과 비지도학습에서 사람이 하는 역할의 차이점을 설명하는 다음 문장의

	이점을 설명해주세요.	빈칸에 들어갈 알맞은 단어를 골라주세요. <문장> 지도학습에서는 사람이 데이터의 [레이블/크기/출처] (을)를 기계에게 알려주어야 한다. 반면, 비지도학습에서는 알려주지 않아도 된다.
8-1	인공지능의 판단이 공정하고 중립적으로 보일 수 있지만 사실은 그렇지 않을 수 있습니다. 그 이유를 인공지능에게 여러분이 학습시킨 단어의 특성을 생각하며 설명하시오.	인공지능의 판단이 공정하고 중립적으로 보일 수 있지만 사실은 그렇지 않을 수 있습니다. 그 이유를 설명한 다음 문장의 2개의 빈칸에 공통적으로 들어갈 알맞은 단어를 답하시오. <문장> 인공지능은 ____(이)가 제공한 데이터를 학습한다. 그런데 불분명한 단어의 경우, ____의 주관이 담기기 때문에 이러한 단어로 학습한 인공지능은 중립적인 판단을 하기 어려울 것이다.
9-1	어떤 용의자가 범죄자인지 아닌지를 얼굴 사진으로 판단해주는 인공지능 기술이 가진 위험성이 무엇인지 배운 내용을 바탕으로 설명하시오.	어떤 용의자가 범죄자인지 아닌지를 얼굴 사진으로 판단해주는 인공지능 기술이 가진 위험성이 무엇인지 설명하는 다음 문장의 빈칸에 들어갈 알맞은 단어(형용사)를 답하시오. <문장> 인공지능이 특정 생김새를 가진 용의자의 이미지 데이터로만 학습한다면, ________ 예측만을 할 것이다.
9-2	인공지능 윤리와 관련된 문제점을 해결하는 바람직한 방법과 그 이유를 설명하시오.	인공지능 윤리와 관련된 문제점을 해결하는 바람직한 방법과 그 이유를 설명한 다음 문장의 빈칸에 들어갈 알맞은 단어 조합을 고르시오.

		<문장> 인공지능으로 인해 발생할 수 있는 다양한 상황들을 살펴보고 [여러 분야의 사람/여러 개발자] 들의 의견을 모아서 기준을 정할 필요가 있다. 왜냐하면, 인공지능이 소수의 사람의 선택만을 학습하면 [객관적인/편향된] 판단을 내릴 수 없기 때문이다.

Appendix #3. Pre and Post-Test (Korean)

[1] 다음 중 기계학습에 대한 알맞은 설명을 하나 골라주세요.

- 기계학습이란 데이터를 사용하여 기계가 의존적으로 학습하게 하는 방법이다.
- 기계학습이란 사람을 사용하여 기계가 의존적으로 학습하게 하는 방법이다.
- 기계학습이란 데이터를 사용하여 기계가 스스로 학습하게 하는 방법이다.
- 기계학습이란 사람을 사용하여 기계가 스스로 학습하게 하는 방법이다.

[2] 사람이 인공지능에게 잘못된 라벨(이름표)이 달린 데이터를 제공할 때 발생할 수 있는 상황에 대해 알맞은 설명을 하나 적어주세요.

[3] 지도학습의 의미에 대한 다음 문장의 빈칸에 들어갈 알맞은 단어를 적어주세요.

지도학습이란 기계학습의 종류 중 하나이며, [ ] (이)가 인공지능에게 [ ] (을)를 지도하여 학습시키는 방법이다.

[4] 지도학습과 비지도학습의 차이에 대한 다음 문장의 빈칸에 들어갈 알맞은 단어를 적어주세요.

지도학습에서는 사람이 [ ] (을)를 기계에게 알려주어야 한다. 반면, 비지도학습에서는 알려주지 않아도 된다.

[5] "편향된 데이터"가 무엇이고 어떤 문제점을 일으키는지 설명해주세요.

[6] 어떤 인공지능 모델을 지도학습 시키는 과정을 설명해주세요.

[7] 인공지능이 학습할 데이터의 품질을 검토할 때 고려해야하는 것들을 2가지 이상 답해주세요.

[8] 의료연구원들은 의료영상 데이터로 기계학습을 학습시켜 질병을 알아보고 진단하도록 학습시킬 수 있습니다. 기계학습이 많은 데이터를 학습할 수 있어도 컴퓨터의 예측에는 문제가 생길 수 있습니다. 어떤 문제점이 발생할 수 있는지 2가지 이상 설명해주세요.

[9] 인공지능 기술을 활용하여 시각장애인이 보지 못하는 물체나 사람 등을 판별하여 음성으로 알려주는 특수안경을 제작하려고 합니다. 이 안경에 쓰일 인공지능 모델을 어떻게 학습시킬지 배운 내용을 바탕으로 설명해주세요.

Abstract in Korean

교육 분야에서는 최근 텍스트 기반 챗봇의 사용이 주목받고 있다. 교육적 목적으로 사용하기 위한 챗봇을 개발하고 디자인할 때 챗봇을 교수자임과 동시에 제품으로 바라볼 수 있다. 따라서 교수자-학습자 상호작용 및 챗봇-사용자 상호작용 측면을 모두 고려해야 한다. 이에 따라, 본 연구를 통해 교수 설계 관련 요인 (학습에서의 인지적 참여) 및 사용자 경험 관련 요인 (챗봇의 휴먼니스의 인지적 결정요인)이 학습 결과와 동기에 미치는 영향을 검토하고자 한다.

본 연구에서는 ICAP 프레임워크와 CMD 프레임워크를 기반으로 네 가지 버전의 챗봇 (Constructive and humanized, constructive and non-humanized, active and humanized, active and non-humanized)이 설계되었다. Constructive 챗봇은 학습자들에게 (1) 비디오/시뮬레이션에서의 개념을 본인의 말로 설명하거나 요약하게 하고, (2) 배운 것에 근거하여 자신의 의견을 정당화하게 하며, (3) 명시적으로 가르친 것에 근거하여 새로운 정보를 추론하게 하고, (4) 학습 중에 다루지 않은 새로운 경우에 대한 예측을 생성하게 한다. 반면에, active 챗봇은 학습자들에게 전체 문장의 공백을 채우거나 다양한 옵션 중에서 정답을 선택하게 하지만, 새로운 출력을 생성하거나 새로운 정보를 추론하도록 요구하지는 않는다. Humanized 챗봇은 인간과 같은 캐릭터의 시각적 표현을 제공하고, (1) 챗봇의 역할과 이름을 인간처럼 식별하고, (2) 사회적 대화를 사용하고, (3) 비공식적이고 캐주얼한 어투를 사용하고, (4) 이모티콘을 사용하고, (5) 얼굴 표정이 있는 스티커를 사용하는 등의 여섯 가지 휴먼니스의 인지적 결정요인을 사용하여 디자인되었다.

본 연구의 첫 번째와 두 번째 가설은 각각 학습의 인지적 참여 모드와 챗봇 인간성의 인지적 결정요인의 주효과를 검증하기 위해 설정되었다. 세 번째 가설은 두 독립 변수의 상호작용 효과를 검증하기 위해 설정되었다. 혼합적 데이터 분석에 근거하여 양적 및 질적 데이터를 종합적으로 분석하여 세 가지 가설을 검증하였다. 사전/사후 테스트를 통해 학습성과를 측정하였고, 사후 설문과 인터뷰를 통해 학습동기를 측정하였다. 측정된 독립변수의 실험군 간 차이가 유의한지 확인하고자 이원 분산 분석과 순위-합 변환 이원 분산 분석을 실시하였다.

본 연구 의의는 네 가지로 요약된다. 첫째, 교수 설계 요소와 사용자 경험 요소를 모두 포괄할 수 있도록 학습 경험 디자인의 개념에 기반하여 교육용 챗봇을 설계하였다. 둘째, 학습자들이 배경 지식을 갖추지 못했음에도 불구하고, constructive 챗봇이 활동적인 챗봇보다 학습 성과를 향상시키는 데 더 효과적이었다. 셋째, 선행연구에 기반한 예측과는 다르게, non-humanized 챗봇이 humanized 챗봇보다 학습 성과 향상에 더 효과적이었다. 넷째, constructive 챗봇을 사용할 때, 챗봇 휴먼니스의 인지적 결정요인은 두 가지 유형의 학습 동기에 대해 다른 영향을 미쳤다. 구체적으로 긴장-압박 측면의 동기에 대해서는 긍정적인 영향을, 자기 유능감 측면의 동기에 대해서는 부정적인 영향을 주었다.