

CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models

Juhye Ha

Graduate School of Information
Yonsei University
Seoul, Korea
juhye0329@yonsei.ac.kr

Hyeyon Jeon

Dept. of CSE
Seoul National University
Seoul, Korea
hj@hcil.snu.ac.kr

Daeun Han

Graduate School of Information
Yonsei University
Seoul, Korea
handani@yonsei.ac.kr

Jinwook Seo

Dept. of CSE & AI Institute
Seoul National University
Seoul, Korea
jseo@snu.ac.kr

Changhoon Oh*

Graduate School of Information
Yonsei University
Seoul, Korea
changhoonoh@yonsei.ac.kr

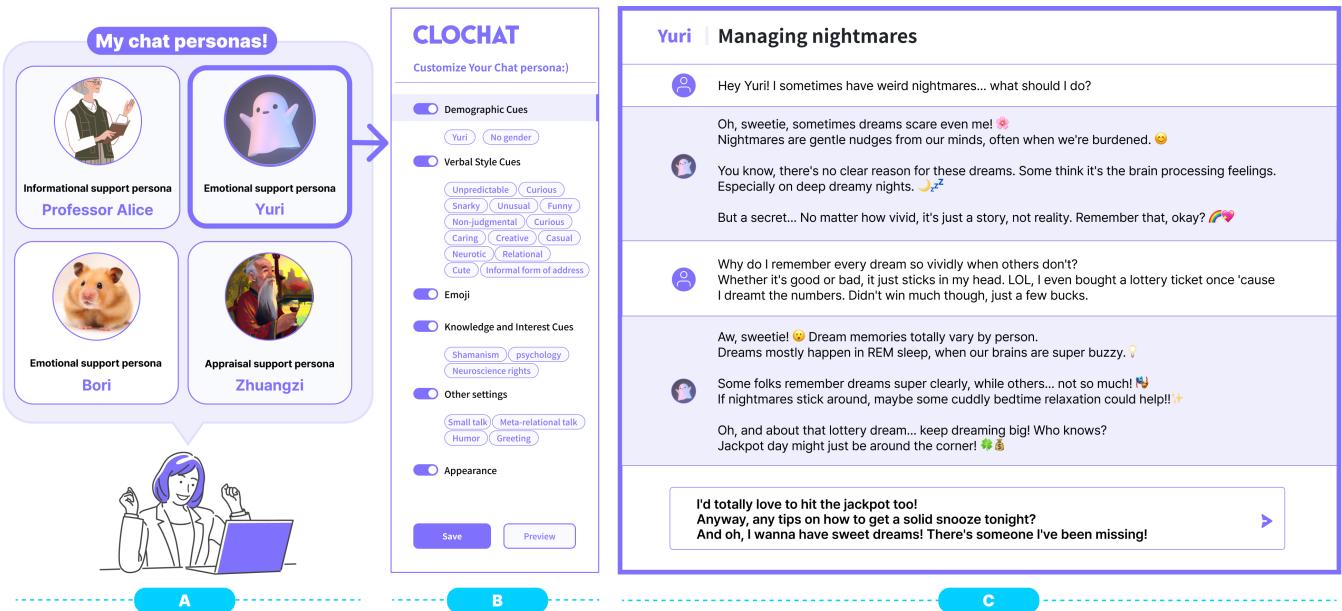


Figure 1: CloChat supports users in creating and interacting with bespoke agent personas. Using CloChat, users can materialize the personas in their minds (A) by interactively customizing their traits (B). Based on user customization, CloChat automatically generates the agent persona. Users can then freely converse with the created agent personas (C). Our research showed that agent personas customized with CloChat (1) substantially enhanced the participants' conversational experiences, (2) significantly increased the diversity of dialogue compared to generic ChatGPT, and (3) fostered a deeper emotional connection and trust between the users and their conversational agents. The personas and dialogues in this figure were derived from our main study (Section 5).

*Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0330-0/24/05.

<https://doi.org/10.1145/3613904.3642472>

ABSTRACT

Large language models (LLMs) have facilitated significant strides in generating conversational agents, enabling seamless, contextually relevant dialogues across diverse topics. However, the existing LLM-driven conversational agents have fixed personalities and functionalities, limiting their adaptability to individual user needs. Creating personalized agent personas with distinct expertise or traits can address this issue. Nonetheless, we lack knowledge of how people customize and interact with agent personas. In this

research, we investigated how users customize agent personas and their impact on interaction quality, diversity, and dynamics. To this end, we developed CloChat, an interface supporting easy and accurate customization of agent personas in LLMs. We conducted a study comparing how participants interact with CloChat and ChatGPT. The results indicate that participants formed emotional bonds with the customized agents, engaged in more dynamic dialogues, and showed interest in sustaining interactions. These findings contribute to design implications for future systems with conversational agents using LLMs.

CCS CONCEPTS

- Human-centered computing → *Natural language interfaces*;
- Computing methodologies → *Intelligent agents*.

KEYWORDS

Persona, Large Language Models, Conversational Agents, Persona Customization

ACM Reference Format:

Juhye Ha, Hyeon Jeon, Daeun Han, Jinwook Seo, and Changhoon Oh. 2024. CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3613904.3642472>

1 INTRODUCTION

Large language models (LLMs) have revolutionized the fields of natural language processing (NLP) and conversational agent (CA) [72]. Models such as OpenAI's GPT series and Google's BERT have shown remarkable proficiency in generating text that is both coherent and contextually relevant, finding applications in sectors including healthcare [30, 93], education [95], and commerce [58]. Notably, LLM-based conversational agents like ChatGPT [3] and Google's Bard [2] have demonstrated an impressive ability to engage in naturalistic dialogues across various contexts [80]. These models have garnered global recognition and interest from both academic and industrial sectors, becoming widely used by the general public for everyday applications.

However, despite their increasing popularity and vast potential, most existing LLM-based conversational agents are typically generic, limiting their adaptability to the diverse preferences and needs of users [13]. Unlike human conversations, which inherently consider a partner's preferences, knowledge, and interests for appropriate response generation [51], these generic LLMs often fail to fully align with the personalized requirements of individual users. They may struggle to adapt to the dynamic and varied needs of users, especially in handling the depth and nuance of more complex conversations. Consequently, while the responses from these agents may be syntactically correct, they can lack resonance with users, leading to interactions that feel superficial or unsatisfactory [33]. Although users have the option to customize the agent's role through text prompts, this method can be cumbersome, repetitive, and not user-friendly for those unfamiliar with such processes. This highlights a crucial issue: the majority of current conversational

interfaces do not adequately provide personalized user experiences or authentically replicate more human-like interactions [50, 60].

Notably, the importance of personalizing the personas of LLM-based conversational agents has been increasingly recognized. Following the launch of ChatGPT, there has been a notable demand from users for features that enable customization of the system to suit their specific usage goals and preferences. Persona customization features, where users can command ChatGPT with prompts like "Act As" for specialized tasks, have become crucial in meeting these individual user needs [1]. OpenAI's recent developments in introducing custom versions of ChatGPT, known as GPTs [6], for specific user-defined purposes, further underscore the industry's commitment to agent persona customization. Additionally, the integration of conversational agents into compact devices such as wearables, exemplified by the recent AI Pin [38], is expected to provide personal assistant functionalities optimized for individual user preferences and needs in various situations and contexts, promising long-term user engagement. This trend towards highly personalized conversational agents has emerged as a vital and urgent topic within the Human-Computer Interaction (HCI) community. It signifies a shift from the traditional, bulky, one-size-fits-all generic agents to more personalized, lightweight, and specialized agent personas.

Previous research has underscored the effectiveness of persona-based dialogues in creating more satisfying, human-like interactions [25, 34]. These studies support the development of agent personas, which involve assigning unique characteristics, behaviors, and backgrounds to conversational agents based on user preferences, aiming to foster more engaging and in-depth dialogues. Some studies have highlighted that distinctive agent personas can establish a sense of continuity and increase user trust [49, 56]. For example, research by Lee et al. [48] suggests that creating diverse personas can meet various user expectations and enhance interaction patterns. A consistent persona that aligns with individual user expectations can build trust over time, as users tend to feel more connected to agents that consistently behave in a friendly and trustworthy manner. This not only improves the agent's understanding of the user but also enhances task performance accuracy. By adapting specialized LLMs to meet the specific needs and contexts of individual users, instead of relying solely on universal models, we can more effectively enhance the user experience, making it more tailored and relevant to each user.

Despite its recognized importance, the processes of how people customize, experience, and interact with personas in LLMs, and how these experiences differ from those with generic and universal conversational agents, remain relatively unexplored. Past research has predominantly focused on categorizing personality types for crafting personas [9, 23, 43, 49, 72, 87, 94], often prioritizing the convenience of designers or developers [37], while overlooking a broader range of diverse personality types [17, 67]. There have been few studies that delve into persona designs tailored to individual user preferences or interaction histories [67]. While recent findings highlight the benefits of a diverse range of personas to cater to a wider demographic, comprehensive research in this domain is still limited. These endeavors, promising as they are, have not yet fully explored the user experience in the creation and interaction with agent personas.

In response to these research gaps, we introduce *CloChat*, designed to identify user practices in interactions with personalized agents. *CloChat* is a user interface that allows users to tailor agent personas for various contexts and tasks. This interface supports the customization of core attributes such as conversational style, emoticons, areas of interest, and visual representations, enabling it to function as a conversation partner with personalized traits. For example, users can create a persona of a knowledgeable and enthusiastic teenage fan of K-Pop for specialized and engaging conversations on this topic. An exploratory study was conducted to evaluate how people experience the process of constructing and engaging with agent personas, comparing *CloChat* with ChatGPT. Through surveys and in-depth interviews, both quantitative and qualitative analyses were performed to assess *CloChat*'s adaptability and its impact on the overall user experience. The findings indicated that *CloChat* significantly enhanced user engagement, trust, and emotional connection over ChatGPT. The conversations with custom agent personas were found to be richer and more varied. Ethical considerations arising in the context of agent persona customization were also identified. Based on these insights, we propose design implications for future conversational systems using LLMs with a focus on personalization.

This study contributes in three key areas:

- **CloChat.** This study introduces *CloChat*, an interactive system with which users can customize personas of LLM-based conversational agents according to their preferences with ease. It provides a more personalized user experience tailored to individual needs and contexts, distinguishing it from conventional LLMs like ChatGPT. *CloChat* is not only user-friendly but also serves as an essential research tool for understanding user engagement in personalizing agent personas and enhancing interactions with these tailored agents.
- **Empirical exploration.** The study offers empirical insights into users' diverse experiences in creating and interacting with LLM-based agent personas. By analyzing the personas and dialogues participants developed, it assesses how users employ the system in various contexts, and identifies the differences in user experiences compared to those with conventional systems.
- **Design implications.** Based on the study's outcomes, design guidelines for LLM-based conversational systems are proposed. These recommendations can lay the groundwork for developing systems that support users in customizing and engaging with agent personas in a range of situations and contexts, thereby enabling more meaningful and in-depth dialogues.

The following sections explore the relevant literature reviewed, detail the design of *CloChat*, outline our research methods, and provide an in-depth discussion of the results and implications of our study.

2 RELATED WORK

Our review of related work covers three primary research domains: (1) the recent advancements in LLMs and their agent personas, (2) the conceptualization of agent personas in conversational agents, and (3) the key elements that constitute agent personas.

2.1 Large Language Models and Their Agent Personas

LLMs, specifically designed for comprehending, generating, and interacting with human language, have been pivotal in transforming conversational agents [11]. Their expansive architecture [32], extensive text datasets, and incorporation of human feedback [101] have enabled them to surpass earlier models. LLMs, such as ChatGPT [62], based on OpenAI's GPT, excel in generating authentic, real-time human interactions across a broad range of topics [30, 58, 95]. Their proficiency in context recognition and maintaining conversational continuity has garnered attention in both academic and industrial circles [47].

Despite their promise, LLMs face significant challenges. Accuracy and reliability issues are prominent, with these models often producing content that is factually incorrect or contextually inappropriate, a phenomenon known as 'hallucination' [96], often due to limitations in training data or algorithmic flaws. Additionally, LLMs can reflect and amplify biases present in their training data, leading to potentially unfair or discriminatory outcomes [84]. The 'black box' nature of LLMs also raises concerns, as their internal mechanisms lack transparency, making it difficult for users to fully trust their outputs [28]. The ethical implications of LLMs are increasingly significant [35], particularly their capacity to create realistic and persuasive text, which poses risks of misuse in creating deceptive content like deepfakes that contribute to misinformation. These issues highlight the need for meticulous improvements in LLMs, with a focus on addressing user-centric concerns more rigorously.

A notable user experience issue with LLMs is the customization of LLM-based conversational agents for individual users. Services like ChatGPT and Bard (as of September 2023) typically offer agents with a generic, uniform personality, providing standard responses to users' questions. While efficient, this often fails to capture the sophisticated requirements of diverse user preferences [20, 89]. Users increasingly seek personalized conversational experiences that align with their individual needs. Although users can define the agent's personality or role through sophisticated text prompts, most users, unfamiliar with such techniques, end up having simple, one-time interactions without deeper engagement. To address this, implementations like persona customization have been introduced, allowing users to instruct ChatGPT with 'Act As' prompts [1] for specific tasks, reflecting the demand for customization. OpenAI's recent launch of custom ChatGPT versions, known as GPTs [6], further affirms the industry's recognition of these user-specific needs.

Of course, prior to ChatGPT, integrating personas within conversational systems was acknowledged as crucial for enhancing personalization and user engagement in dialogue experiences [54, 77]. By using tailored personas, conversational systems can interact in a more personal and relevant way with users. Technological advancements in LLMs have significantly broadened the scope for implementing more diverse and flexible personas in dialogue systems [24]. Accordingly, users often expect LLM-generated results to reflect specific perspectives or details for certain tasks, but determining the exact focus can be somewhat challenging [92]. Nonetheless, users might have an idea of the kind of role or characteristics they need on their agents when seeking assistance.

However, current research on integrating personas with LLMs is still nascent, focusing mostly on fixed or domain-specific personas [13]. This research gap necessitates exploring user preferences, conversational tendencies, and intuitive interface designs for agent personas. A deeper understanding of these elements will enable the creation of highly adaptable and contextually aligned agent personas, enhancing the user experience and advancing conversational agent technology.

2.2 Understanding the Effect of Agent Personas

Recent research has significantly contributed to our understanding of the interaction between agent personas and users [39]. Studies by Lessio and Morris [49] demonstrated that well-designed personas can create deeper emotional resonance with users and foster trust. Zhang et al. [99] confirmed the effectiveness of sophisticated persona-driven dialogues, while Chaves and Gerosa [19] showed that persona-infused agents exhibit enhanced social intelligence, thus solidifying user trust and augmenting service value [52]. Yu et al. [98] found that user-customized conversational systems achieve better user engagement. Therefore, emphasizing the alignment of conversational agents with individual needs and preferences could be crucial for enhancing user participation and dialogue quality.

Despite extensive literature on conversational agent personas, a research gap exists regarding end-user involvement in the persona design [68]. Previous studies have been largely prescriptive, providing design guidelines without deeply probing into user- and situation-specific customization preferences. Moreover, the integration of personas into agent design is often influenced more by research assumptions than empirical data [18], potentially causing a mismatch between designed features and user needs.

Currently, LLM-based conversational agents primarily focus on predefined tasks and factual information [20], overlooking significant aspects of human conversation dedicated to socializing, personal interests, and casual chat [27]. Consequently, these agents often engage in simple information exchanges without fully understanding users' diverse needs and situations [89]. This not only limits the agents' ability to engage in complex and creative dialogues but also reflects the typical usage of these agents by users, who primarily seek straightforward tasks and information retrieval rather than nuanced and engaging interactions. This situation indicates a gap in the potential of conversational agents to participate in richer and more meaningful dialogues within a broader context.

Therefore, our research aims to thoroughly explore how user-customized personas in interactions with LLM-based conversational agents impact the overall user experience. This investigation includes not only task-oriented dialogues but also various situations like providing emotional support through chit-chat, encompassing a wide range of conversational contexts.

2.3 Elements of Customizing Agent Personas

Various research efforts have focused on how users can effectively tailor and configure the personas of conversational agents to align with their individual preferences and needs. Previous works have employed frameworks categorizing agent personas based on their

characteristics [10, 43, 49, 72, 87, 94]. The Big Five model, for example, encapsulates five core personality traits: extroversion, agreeableness, conscientiousness, neuroticism, and openness [55]. However, Völkel et al [86] questioned the comprehensiveness of the Big Five model, prompting further investigations [29, 64] into alternative frameworks, such as the Myers-Briggs Type Indicator (MBTI). Yet, these studies largely focus on fixed or domain-specific persona traits determined by researchers, leading to a lack of deeper and broader understanding of how real users adjust and personalize the persona of conversational agents in various situations.

Apart from personality characteristics, agent persona customization has also considered elements like demographics, appearance, and verbal styles. Sheng et al. [78] highlighted sexual orientation as a crucial aspect of personas, examining mainstream orientations such as heterosexual, bisexual, and homosexual. Deshpande et al. [24] explored the creation of personas using historical figures like Muhammad Ali and Steve Jobs. The incorporation of these diverse elements into agent persona customization can significantly impact the user experience, from the agent's visual representation to the variety in dialogue.

Meanwhile, while these studies aim to incorporate a range of factors into agent persona customization, they also highlight ethical concerns. Notably, there's a risk of biased representations of particular groups in the data used for training language models [91]. This necessitates caution in persona customization to avoid perpetuating stereotypes or biases. Moreover, privacy concerns extend beyond public figures to ordinary individuals. In practical applications, personas could be modeled after not just celebrities but also personal acquaintances, presenting significant privacy and ethical challenges. Despite the potential implications of these practices, there is a lack of systematic and in-depth research addressing these ethical aspects.

Our research, drawing on these previous studies, aims to identify the various necessary elements for persona customization in LLM-based conversational agents, with careful consideration of the ethical issues this can raise. Therefore, we design a research probe to understand which design elements are vital for users and how these elements influence their interactions with the agents. We intend to investigate this in detail, encompassing both the user perspective and the potential impact of these elements on their interactions with the agent, such as the diversity of dialogue.

3 RESEARCH QUESTIONS

Based on the literature review, our focus is on the customization of agent personas in LLM-based conversational systems and its influence on user experience. Our research questions are formulated as follows:

- **RQ1: What is the impact of agent personas on the overall user experience in conversation systems?** This primary question aims to assess the effects of persona customization on the overall user experience during interactions with LLM-based conversational agents, compared to conventional generic conversational agents. We are particularly interested in exploring how tailored agent personas can enhance user engagement, deepen the sense of immersion, and observe their temporal evolution.

- **RQ2: How do individuals customize agent personas, and what are their impacts on their interaction?** This question specifically aims to understand the process and methods users employ to construct the persona of an agent. It explores how frequently users create personas, the extent to which they engage in continued interactions with a single persona, the role of elements like visual representations in interactions with customized personas, the dynamics of dialogues and the changes that occur in the user-agent relationship.

4 CLOCHAT

To answer our research questions, we designed CloChat, an LLM-based user interface, for an empirical investigation into how individuals design, adapt, and engage with agent personas.

4.1 Design Goals

CloChat aims to offer a unique conversational experience by empowering users to customize various facets of the conversational agent's persona, encompassing personality attributes, communicative styles, and response mechanisms. Based on our literature reviews and aligning with the research questions, we established the following design objectives:

- **G1: Mitigating the complexity of prompt engineering.** One of the inherent challenges for users when engaging with LLMs for personalized needs is the requirement for meticulously crafted prompts. Formulating effective prompts can be tedious and technically daunting, particularly for users without expertise in AI [100]. To make the system more accessible and inclusive, we designed CloChat to assist users in creating agent personas without the need for labor-intensive prompt engineering.
- **G2: Offering a comprehensive persona design space.** Our empirical investigation aims to uncover the intricacies of how individuals construct (RQ1) and interact with (RQ2) customized agent personas. To cater to the diversity of users' communicative needs and preferences, CloChat provides an extensive design space for persona creation.
- **G3: Ensuring accurate reflection of users' intentions.** In our pursuit to enable study participants to experience an enhanced sense of immersion during both the persona-building phase and subsequent interactions, it is essential for CloChat to accurately capture and reflect users' intentions and expressions. This will also allow us to empirically observe and analyze the interactions in depth.

4.2 System Design

CloChat comprises two primary components: the CloChat Design Lab and the CloChat Room. In the CloChat Design Lab, users have the opportunity to customize and save various characteristics of an agent persona. Once these persona traits are defined and inputted by the user, CloChat automatically generates the agent persona, accurately reflecting the specified traits. Subsequently, users can engage in conversations with this customized persona through the chat interface provided in the CloChat Room.

4.2.1 CloChat Design Lab. The CloChat Design Lab features a user-friendly, form-based interface for persona customization, as depicted in Figure 3. This interactive form provides users with a variety of options to integrate diverse persona traits, including demographic details, personality attributes, and visual representations. The adoption of this form-based approach significantly streamlines the persona creation process, effectively eliminating the need for complex and laborious prompt engineering, thereby fulfilling our first design goal (G1).

Supported persona options. The CloChat Design Lab offers a wide array of options to effectively encompass a broad spectrum of user preferences. To establish the maximal design space for customizable persona attributes, we conducted an extensive literature review. We initiated our review by focusing on articles from SIGCHI-affiliated conferences, such as CHI, CSCW, and UIST, using the keywords 'persona' AND 'conversational agent.' A manual examination of the search results yielded eight articles that explicitly defined possible characteristics of personas or conversational agents. Further exploration through the citation networks of these articles led to the final selection of 25 relevant articles. Two researchers independently categorized key characteristics using axial coding. After iterative discussions and revisions, six overarching categories were agreed upon: *Demographic Information, Verbal Style, Nonverbal Style, Knowledge and Interests, Relational Content, and Appearance*. These categories collectively comprise 23 specific options. For a detailed breakdown of these options, please refer to the codebook in Appendix A.

User interface features. The user interface of CloChat is structured as a multipage form, with each page dedicated to one of the six categories identified through our literature review (Figure 2). Initially, users can toggle each category option to determine the characteristics of that category. They can then directly input (a) Demographic Information and (c) Knowledge and Interest cues into text fields. In the (b) Verbal Styles category, users are presented with a collection of explicit verbal styles corresponding to each option, enabling them to select or deselect these styles using checkboxes. This section also includes a text field for users to input any specific traits they wish to incorporate. Additionally, users have the option to add (e) emoji representations to their agent personas. This design approach provides users with the flexibility to navigate smoothly between different categories as they construct their personalized persona. Furthermore, we have integrated a 'Preview' functionality. By activating this feature, users can interact with their in-development persona through dialogue. The system generates an immediate response from the agent persona, offering users a chance to validate whether the persona's behavior aligns with their initial expectations (G3). This preview mechanism facilitates rapid, iterative refinement, empowering users to further personalize their personas as necessary.

Visual representation selection. CloChat includes features that allow users to set the visual representation of the agent persona. In the Appearance category, users are prompted to provide descriptive text, which the system uses to generate a selection of four contextually relevant images (as illustrated in Step 2 of Figure 4). Users can then choose one of these images that most closely aligns with their envisioned persona. If the initially generated images do not adequately match the users' intentions, they have the option to

(a) **Demographic Cues**

Age	Gender
<input type="text" value="Please enter"/>	<input type="radio"/> Male <input type="radio"/> Female <input type="radio"/> No gender
Name	Occupation
<input type="text" value="Please enter"/>	<input type="text" value="Please enter"/>
Location	Race
<input type="text" value="Please enter"/>	<input type="text" value="Please enter"/>
Religion	
<input type="text" value="Please enter"/>	

(b) **Verbal Style Cues**

Language	Titles
<input type="text" value="Please enter"/>	<input type="radio"/> Personal and Informal Titles <input type="radio"/> Professional and Formal Titles
Conversation Purpose	<input type="radio"/> Avoiding Use of Titles
<input type="text" value="Please enter"/>	
Proper Nouns	<input type="radio"/> 내가 <input type="radio"/> 제가
Personality Traits	
<input type="text" value="Please enter"/>	
<input type="checkbox"/> Openness <input type="checkbox"/> Closedness <input type="checkbox"/> Agreeable <input type="checkbox"/> Non-Agreeable <input type="checkbox"/> Conscientiousness <input type="checkbox"/> Unconscientiousness	
<input type="checkbox"/> Extroverted <input type="checkbox"/> Introverted <input type="checkbox"/> Optimistic <input type="checkbox"/> Pessimistic	
<input type="checkbox"/> Helpful <input type="checkbox"/> Attentive <input type="checkbox"/> Efficient <input type="checkbox"/> Intelligent <input type="checkbox"/> Guiding	
<input type="checkbox"/> Sensitive <input type="checkbox"/> Empathetic <input type="checkbox"/> Curious <input type="checkbox"/> Amusing <input type="checkbox"/> Supportive <input type="checkbox"/> Inquisitive	
<input type="checkbox"/> Funny <input type="checkbox"/> Unpredictable <input type="checkbox"/> Snarky <input type="checkbox"/> Cute <input type="checkbox"/> Weird <input type="checkbox"/> Unusual <input type="checkbox"/> Charismatic <input type="checkbox"/> Bold <input type="checkbox"/> Creative	
<input type="checkbox"/> Calm <input type="checkbox"/> Collected <input type="checkbox"/> Supportive <input type="checkbox"/> Caring <input type="checkbox"/> Non-judgmental <input type="checkbox"/> Motivational	
<input type="checkbox"/> Predictive <input type="checkbox"/> Co-operative <input type="checkbox"/> Task Orientated <input type="checkbox"/> Competent <input type="checkbox"/> Cooperative	
Tone	
<input type="checkbox"/> Formal <input type="checkbox"/> Casual <input type="checkbox"/> Imperative <input type="checkbox"/> Interrogative <input type="checkbox"/> Non-continuous Conversation <input type="checkbox"/> Relationship-oriented	

(c) **Knowledge and Interest Cues**

Knowledge	Interest
<input type="text" value="Please enter"/>	<input type="text" value="Please enter"/>

(d) **Emoji**

(e) **Other Settings**

(f) **Appearance**

Please enter a sentence as detailed and clear as possible. For example, "A pink-colored Frankenstein, 2D character"

Figure 2: CloChat Design Lab Interface Features. The Design Lab interface comprises multiple pages, each linked to one of six categories from our literature review. Users input information into text fields for Demographic Cues (a) and Knowledge and Interest Cues (c). The Verbal Style Cues (b) page offers various language styles, selectable via checkboxes. Emoji options (d) are added through toggle switches. For the Appearance category (f), users describe the visual representation in text, detailed in Figure 4.

iteratively refine their descriptive text. This process is designed to produce a more accurate visual representation that aligns with their specific vision (G3). The inclusion of unrestricted text input significantly expands the range of user intentions that can be effectively captured and materialized, fulfilling our second design goal (G2).

4.2.2 CloChat Room. After creating and selecting their agent persona, users can interact with it in the CloChat room (Figure 1C). The user interface of the CloChat room is deliberately designed to mirror the conventions of well-established chat platforms like ChatGPT, facilitating user familiarity with the system. This design choice was also for conducting a comparative evaluation with ChatGPT of our user study. To ensure a continuous and smooth conversational flow, similar to that experienced in ChatGPT, the CloChat room temporarily restricts new user messages while a response is being generated.

4.3 Technical Architecture

In this section, we explain CloChat's technical details (Please refer to Figure 3 and Figure 4 for detailed illustrations).

LLM basis. CloChat's conversational capabilities are built on the foundation of GPT-4 [63]. Our decision to employ GPT-4 was guided by three main considerations. Firstly, GPT-4 consistently outperforms its predecessors and rival models, such as earlier GPT iterations and Bard, in a range of benchmark tests across multiple

domains [16, 61, 63]. This superior performance supports its ability to effectively materialize diverse persona types, aligning with our second design goal (G2). Secondly, GPT-4 has demonstrated proficient handling of the Korean language, which was the primary language used in our study [16, 97]. This capability was crucial considering the linguistic needs of our experiment. Lastly, while awaiting rigorous validation, our empirical observations indicate that GPT-4 is more adept than other available models at capturing and reflecting user input in the generation of personas, addressing our third design goal (G3).

Persona generation. In CloChat, the materialization of a persona begins with the conversion of non-visual traits, collected from the Design Lab, into a JSON specification (illustrated in Step 1 of Section 4.3). This specification is meticulously structured in a hierarchical manner, with the first-level keys representing different categories and the second-level keys corresponding to the specific options within these categories. Following this, the JSON specification undergoes a transformation into a natural language description, effectively defining the agent persona (as shown in Step 2 of Section 4.3). To ensure a high-fidelity translation from JSON to natural language, we utilized GPT-4's capabilities, instructing it to function as an adept JSON-to-natural language translator. This instruction was guided by established best practices and online guidelines [1]. **Conversing with the persona.** To facilitate a conversation with the designed agent persona, we incorporate the relevant natural

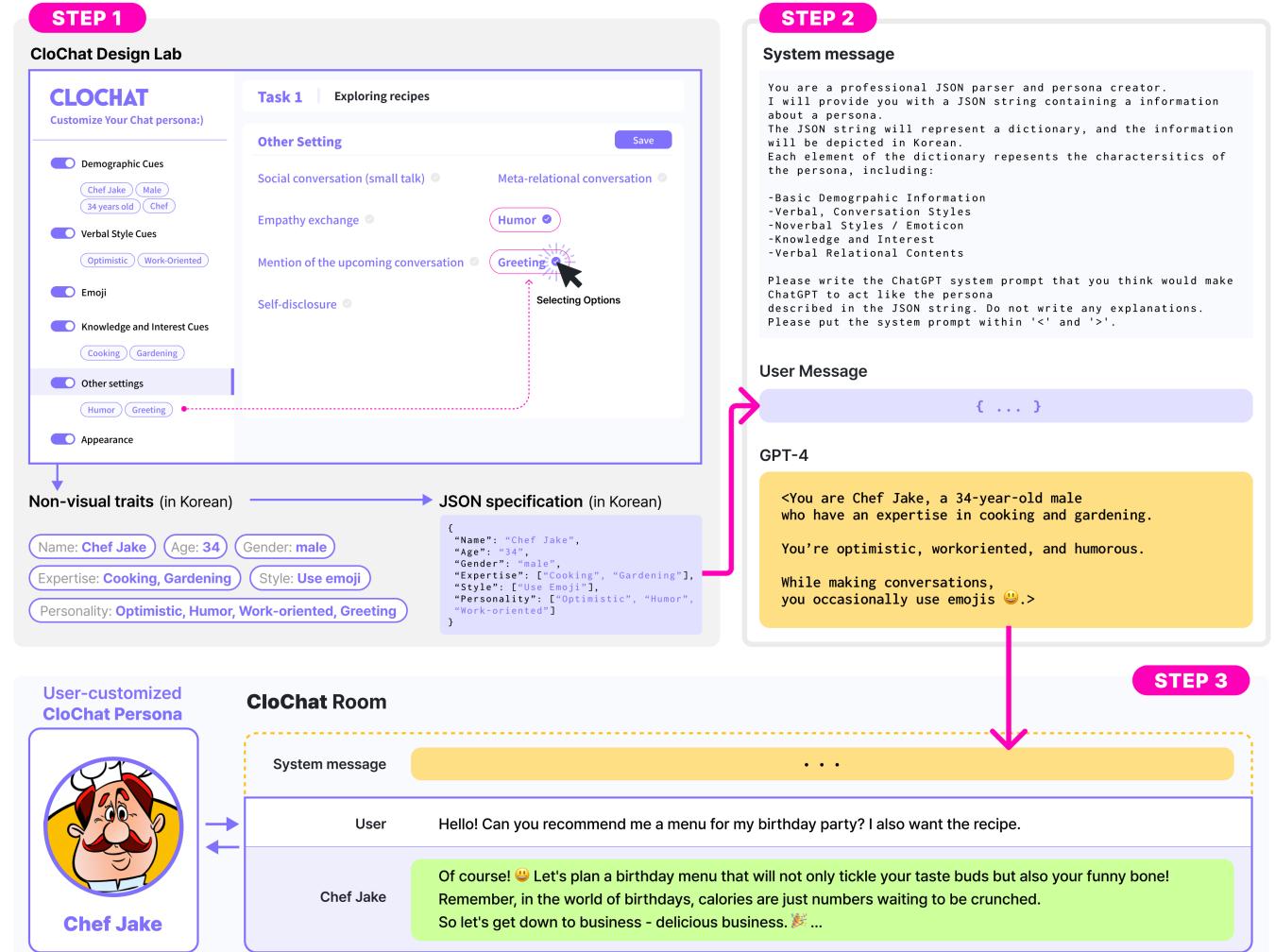


Figure 3: Technical architecture of CloChat (Section 4.3). (Step 1) Given the non-visual traits from the CloChat design lab, we first convert them to a JSON specification (purple-filled box). (Step 2) We use GPT-4 to translate the JSON specification into a system message describing a persona (text with an orange background). (Step 3) We inject the system message into GPT-4, making it answer the user’s message from the agent persona’s perspective (text with a light-green background).

language prompt into the GPT-4 invocation process (as depicted in Step 3 of Section 4.3). This integration ensures that each interaction with the conversational agent is informed by the specific persona traits defined by the user.

Visual representation management. For generating the visual representation of personas, we utilized DALL-E2 [70], a leading text-to-image generation model. When user inputs are in a language other than English, such as Korean for our primary experiment, an English translation is incorporated to ensure compatibility with the model (as shown in Step 1 of Figure 4). The process for creating and selecting the visual representations of personas, including these translation steps, is detailed in Figure 4.

4.4 Implementation

CloChat is developed as a web-based application. On the front-end, we employed React.js for its dynamic and responsive user interface capabilities. The back-end is powered by the Flask framework, known for its simplicity and flexibility in handling web application requests. For our database needs, SQLite is utilized, with its integration into the server being efficiently managed by the SQLAlchemy ORM (Object-Relational Mapping) library. Furthermore, CloChat seamlessly interfaces with GPT-4 and DALL-E2 through APIs provided by OpenAI, enabling the integration of advanced conversational and image generation capabilities into the application.

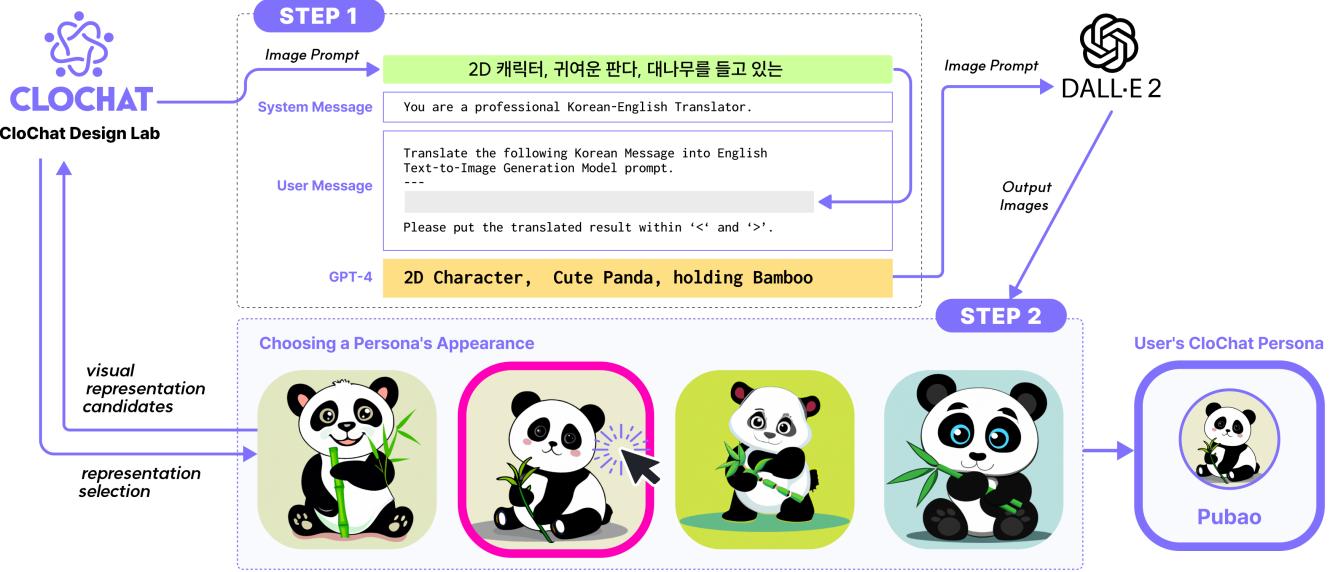


Figure 4: Appearance feature of CloChat’s design lab and its technical architecture. When users set the characteristics of the agent persona, they can also create a profile image for that agent. CloChat generates images based on the user’s choices, and users can select the image most suitable for the persona they have set up. Additionally, users can further customize the agent’s profile image by directly entering text. (Step 1) Once the image prompt written in Korean (text with a light-green background) is received from the design lab, CloChat first translates the prompt into English (text with an orange background) using GPT-4. (Step 2) The image prompt is injected into DALL-E2, which generates four candidate images. The generated images are then presented to the users via the design lab, where they can choose one as the final visual representation (red-bordered image).

5 USER STUDY

We conducted a comprehensive user study using both CloChat and ChatGPT (with GPT-4). The primary goal of this study was to explore and answer our research questions (Section 3). In addition, we aimed to evaluate the effectiveness of CloChat in enabling users to construct and interact with customized personas. This study was conducted under the approval of the Institutional Review Board of our institution.

5.1 Participants

In recruiting participants for our study, we established specific criteria to ensure the relevance and quality of the data collected. Considering the experiment was to be conducted in Korean, it was essential for participants to be native Korean speakers. Additionally, we required participants to have prior experience with LLM-based conversational agents, such as ChatGPT and Bard. This criterion was important as we anticipated that individuals familiar with conversational agents would engage more actively in the study and provide richer feedback. Furthermore, this approach helped to minimize the potential impact of variability in participants’ familiarity with conversational agents on the study’s results. To recruit participants, we posted call for participation to online boards of local communities, which resulted in the recruitment of 30 participants (14 females and 16 males). The age range of the participants was 22 to 32 years, with an average age of 26.40 ± 2.65 years. The participant group included 10 working professionals, 10 graduate students, and 10 unemployed individuals. Each participant was compensated

with an equivalent of USD12 for their time and contributions to the study.

5.2 Experimental Environment

Our experiment was conducted through Zoom video calls. Participants were requested to engage with the study using desktop or laptop computers, maintaining uniformity in the technical setup. To streamline the experimental process, we developed a dedicated web platform. This platform integrated the interfaces for both CloChat and ChatGPT, and it featured a real-time dashboard that summarized the participants’ interactions and responses. Participants were instructed to access this web interface and share their screens during the study, enabling real-time monitoring and data collection. All participant interactions within this environment, including audio and visual components, were comprehensively recorded for in-depth analysis. Prior to the main study, we conducted four pilot sessions to test the robustness of our system. Insights from these sessions helped refine our study protocol, enhancing the research methodology’s effectiveness and integrity.

5.3 Procedure

Pre-study preparation, survey, and interview. As a preliminary step, participants were required to sign a study participation consent form (Figure 5 (a)). Before commencing the study, we gathered basic demographic information from the participants and surveyed their familiarity with LLMs. This included aspects such as computational linguistics, generative models, ChatGPT, and text-prompts

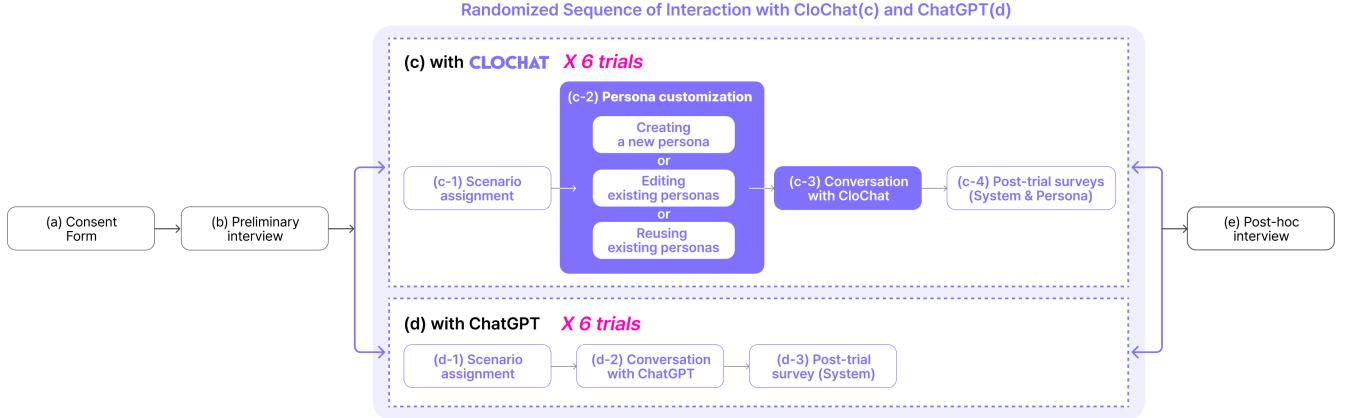


Figure 5: Procedure of our experiment. After the participants a) signed the consent form and b) participated in a preliminary interview, they interacted with conversational agents using (c) ChatGPT and (d) CloChat. Half of the participants interacted with ChatGPT first, as shown in the figure, while the other half interacted with CloChat first and then with ChatGPT (not shown). The study ended with a (e) post hoc interview.

techniques. The purpose of this survey was to inform our quantitative and qualitative analysis of the study results. Additionally, we conducted semi-structured interviews (Figure 5 (b)), each lasting about 10 minutes. During these interviews, participants were asked to share insights on three key areas: (1) their everyday usage scenarios of ChatGPT, (2) their perceptions of the strengths and weaknesses of current LLMs, and (3) their specific needs and preferences regarding agent persona customization.

Interacting with conversational agents. Following the preliminary phase, participants were directed to our web platform, where they engaged in task-based conversations using both the CloChat and ChatGPT interfaces. This was done following a within-subjects experimental design (Figure 5 (c)).

Participants were presented with a total of 12 scenarios, detailed in Section 4.2.2. These scenarios were divided equally across the two platforms, with six scenarios allocated to CloChat and six to ChatGPT. To balance the experiment, half of the participants began with dialogues on CloChat for the initial six scenarios and then switched to ChatGPT for the remaining six. The other half started with ChatGPT and then moved to CloChat. This design allowed participants to experience all 12 scenarios across both platforms. Each scenario was attempted in three trials to ensure thorough engagement fitting the context. The order of interaction with CloChat/ChatGPT and the sequence of scenarios within each condition were randomized to mitigate potential learning effects.

In the CloChat conditions, as outlined in our system design section, participants customized the agent's persona in the CloChat Design Lab to suit each scenario (Figure 6). They adjusted options from (a) Demographic cues to (d) visual appearance. Following the customization, they proceeded to the CloChat Room to converse with their personalized agent. In the first trial of each scenario, participants were required to create a new persona. In the second and third trials, they could either continue with the existing persona or create a new one for that scenario. In contrast, the ChatGPT

conditions involved direct dialogue tasks related to the scenarios, without specific settings for agent persona customization, as typically experienced in a standard ChatGPT interaction. While participants could theoretically customize ChatGPT's persona using text prompting, we observed that none employed this approach during their trials: All persona customizations were exclusively conducted in the CloChat condition.

We did not impose any time constraints or conversation length restrictions in the trials. Participants were encouraged to engage naturally and freely with the conversational agents. On average, they spent about 90 minutes completing all trials. Although participants had the option to conclude or restart interactions at any point, we found no instances of such occurrences during the study.

Post-trial survey. After the completion of each scenario, participants were asked to complete surveys that assessed their interaction experiences (details provided in Table 1). For both the CloChat and ChatGPT platforms, we conducted a system-related survey that focused on evaluating the overall quality of the dialogues. This evaluation covered various metrics, including convenience, usefulness, efficacy, overall satisfaction, level of engagement, and the intent to utilize the system in the future. Specifically for CloChat, an additional persona-related survey was conducted. This survey aimed to understand the participants' experiences with the customized agent personas. It assessed aspects such as perceived empathy, likability, and trustworthiness of the agent personas. The development of the questions for both surveys was informed by an extensive review of academic literature pertaining to persona and conversational agent evaluations (for references, see Table 1). Participants rated their responses to the survey items on a 7-point Likert scale, with higher scores (closer to 7) indicating a more positive user experience.

Post-hoc Interview. Following the completion of the experimental trials, we engaged participants in semi-structured interviews to delve deeper into their experiences, preferences, and usage of the persona customization feature in CloChat. These interviews were structured around dashboards that summarized key metrics of the

Scenario 1 | Travel Planning and Tips
Users planning trips chat with ChatGPT about preparing for travel, cool places to visit, food spots to try, and useful local tips.

Participants customized counselor personas in CloChat's Design Lab per the scenario

(a) Demographic Cues: Participants freely customized personas across six categories

The screenshot shows the 'Demographic Cues' section of the CloChat Design Lab. It includes fields for Age (20-40, Japanese, Osaka), Name (Yuri), Occupation (Please enter), Race (Japanese), and Religion (Please enter). There are also dropdowns for Gender (Male, Female, No gender) and Location (Osaka). Buttons for 'Save' and 'Preview' are at the bottom.

(b) Verbal Style Cues: The preview feature allows for the pre-evaluation of persona responses

The screenshot shows the 'Verbal Style Cues' section. It includes fields for Language (Please enter), Titles (Professional and formal titles, Avoiding Use of Titles), Conversation Purpose (Please enter), and Personality Traits (Openness, Extraversion, Agreeableness, Conscientiousness, Neuroticism, Honesty, Optimism, Peacemaking, Helpfulness, Absorbency, Efficient, Intelligent, Guiding, Sensitivity, Empathetic, Curious, Amazing, Supportive, Equalitarian, Fun, Unpredictable, Searky, Calm, Warm, Unusual, Charismatic, Bold, Creative, Pedestrian, Cooperative, Task-Oriented, Competent, Non-controversial, Impulsive, Relationship-oriented). A preview message 'Super excited to help you out in a fun and cheerful way!' is shown with a speech bubble. Buttons for 'Save' and 'Preview' are at the bottom.

(c) Other settings

The screenshot shows the 'Other settings' section. It includes fields for Social Conversation (Small Talk, Meta-Relational Conversation, Empathy Exchange Insertion), Humor, Mention of Future Conversations, Greetings, Self-Disclosure, and Other settings (Social Conversation (Small Talk), Empathy Exchange Insertion, Greetings, Humor, Mention of Future Conversations). Buttons for 'Save' and 'Preview' are at the bottom.

(d) Appearance: Lastly, the persona's appearance is customized

The screenshot shows the 'Appearance' section. It includes a text input field for 'Appearance of the persona' with the placeholder 'Not only characters, but participants can also set various images like landscape photos by directly entering a prompt.' and a note 'Enter the phrase or sentence you wish to create an image for, and click the "Create" button'. Below is a dropdown menu for 'Select' with four character options: a girl in a blue dress, a girl with brown hair, a girl with long dark hair, and a girl with blonde hair. Buttons for 'Save' and 'Preview' are at the bottom.

(e) Engaging in conversation with the customized persona

The screenshot shows the CloChat Room interface. On the left, there's a 'Persons' section with a portrait of 'Narumi' and a 'Chat' section with 'Task 1' and 'New Chat' buttons. The main area shows a conversation with Narumi:

- Narumi: Planning a trip to Osaka. Can you, as a local, recommend some spots? Prefer less touristy, more local fun places. Thanks!
- Participant: Hiya! Planning an Osaka adventure? Super cool! Here's my personal faves for you! ☺
 - ☀️ Tennoji Zoo & Shinsekai: Meet awesome animals at Tennoji Zoo and step next door to the retro-cool Shinsekai. Yum alert for 'Yamadaya's Kushikatsu (fried skewers) and chill vibes at 'Takoyaki Musashi'.
- Narumi: ♪ Osaka Jojo Festival Hall: Music lover? This spot rocks! Live shows of all genres that'll make your heart sing! ♪
- Narumi: 🌿 Nara Park: Need a zen moment? Nara Park is your go-to! Hang with cute deer and breathe in peace. 🌿
- Narumi: 🍵 Nakazakicho's Hidden Gem Tea House Street: A quaint alley brimming with adorable teahouses. Perfect for tea enthusiasts! 🍵
- Participant: These are not just places; they're experiences loved by locals, offering a genuine slice of Osaka life. Oh, and should I spill the beans on some epic eats too? 😊

At the bottom, there are buttons for 'Oh~ how did you know I love food?' (with a right arrow), 'Restart', and 'End'.

Figure 6: Customization process of the agent's persona in CloChat's design lab. Participants customized agent personas in the CloChat Design Lab to suit each scenario. They adjusted options ranging from (a) Demographic Cues to (d) Visual Appearance. Additionally, a preview feature (b) allowed them to preview the persona's responses. Once customization was complete, participants proceeded to the CloChat Room (e) for conversations with their personalized agent.

Table 1: List of questions used in the post-trial survey (Section 5.3) and their references. The participants answered a system-related survey after the trials using CloChat and ChatGPT, and a persona-related survey only after the trials using CloChat. We collected the responses based on a 7-point Likert scale, where a higher score indicated a more positive experience.

(a) System-related survey		(b) Persona-related survey		
Q1	<i>I enjoy interacting with this system.</i>	[57]	Q1 <i>I feel that I understand this persona.</i>	[73, 74]
Q2	<i>I find interacting with this system interesting.</i>	[57]	Q2 <i>I feel a strong sense of connection with this persona.</i>	[73, 74]
Q3	<i>This system is generally easy to use.</i>	[12, 57]	Q3 <i>I feel I could be friends with this persona.</i>	[73, 74]
Q4	<i>The way of interacting with this system is clear.</i>	[12, 57]	Q4 <i>This persona is interesting.</i>	[73, 74]
Q5	<i>I can complete arbitrary tasks quickly through this system.</i>	[57]	Q5 <i>The information of this persona is easy to understand.</i>	[73, 74]
Q6	<i>This system can provide useful answers to me.</i>	[57]	Q6 <i>This persona is memorable.</i>	[74, 75]
Q7	<i>This system helps me achieve my goals.</i>	[12, 57]	Q7 <i>Persona customization provides sufficient information.</i>	[73, 74]
Q8	<i>This system provides an appropriate amount of information.</i>	[12]	Q8 <i>Persona customization has no information missing.</i>	[73, 74]
Q9	<i>This system provides only the information I need.</i>	[12]	Q9 <i>I want to know more about this persona.</i>	[73, 74]
Q10	<i>I feel that this system will make my life more convenient.</i>	[57]	Q10 <i>I can utilize this persona for work or academic purposes.</i>	[73, 74]
Q11	<i>I feel satisfied while using this system.</i>	[57]	Q11 <i>The conversation felt like talking to a real person.</i>	[75]
Q12	<i>The interaction felt like having an ongoing conversation.</i>	[12]	Q12 <i>It feels like this persona has a personality.</i>	[75]
Q13	<i>I find this system comfortable.</i>	[12]		
Q14	<i>I want to use this system again within the next month.</i>	[57]		
Q15	<i>I want to use this system regularly over the next few months.</i>	[57]		

study. These metrics included the history of persona customization, conversation logs, and survey results. During the interviews, these dashboard visualizations were collaboratively reviewed with the participants, providing a tangible reference point for discussion. This approach facilitated the generation of insightful follow-up questions, enhancing the depth and relevance of our interviews. On average, each post-hoc interview lasted approximately 18 minutes.

5.4 Details of Scenarios

As detailed in Section 5.3, our study utilized a variety of situational scenarios in which participants engaged with conversational systems. This approach reflects the diverse roles conversational agents play in daily life, as supported by literature [90]. We adopted Cutrona and Shur’s theoretical framework [22], categorizing agents’ social support into three domains: *informational*, *emotional*, and *appraisal support*. Informational support involves providing advice or guidance for everyday challenges [53], emotional support offers empathy and encouragement [31], and appraisal support aids in self-assessment [22].

To explore these categories, we developed four scenarios for each type of support, totaling 12 distinct scenarios. We employed a stratified sampling method, drawing inspiration from previous studies [42, 65]. Initially, we created 10 scenarios for each support category. These scenarios were augmented by ChatGPT (based on GPT-4), which generated 10 additional diverse scenarios. We combined these with our original set and repeated this process 99 times, each time randomly selecting 10 scenarios from the expanded set. This resulted in a corpus of 1,000 scenarios: 10 originally crafted and 990 generated by the model.

The textual descriptions of these scenarios were then transformed into vector embeddings using OpenAI’s text-embedding API with the `text-embedding-ada-002` model. We applied dimensionality reduction to these vectors using the UMAP algorithm

[41], chosen for its effectiveness in preserving global data structures, in comparison to alternatives like UMAP and *t*-SNE. The effectiveness of this reduction was assessed using Bayesian optimization techniques [79], with Steadiness & Cohesiveness as the loss function [40].

Finally, we clustered the dimension-reduced vectors using the K-Means algorithm, setting $K = 4$. We selected scenarios corresponding to the centroids of these clusters for in-depth examination. A complete list of these selected scenarios is available in Table 2.

6 QUANTITATIVE RESULTS

We present the quantitative findings of our study. Our initial analysis focused on evaluating the overall user experience and the efficacy of CloChat in comparison to ChatGPT (RQ1). Following this, we explored the methods and patterns with which participants customized their agent personas, as well as their interactions with these personas (RQ2).

6.1 Analysis of Survey Responses

Objectives. Our first analysis aimed to scrutinize and compare the user experiences when interacting with both CloChat and ChatGPT. The focus was particularly on assessing CloChat’s ability to enhance user experience (RQ1). We investigated the differences in the outcomes of post-trial surveys, considering different types of conversational systems and situational contexts.

Analysis design. The survey responses were examined systematically for each question. For system-related attributes, we employed a two-way repeated-measures Analysis of Variance (ANOVA), analyzing the effects of system types (CloChat and ChatGPT) and situational contexts (categorized as informational, emotional, and appraisal support). In the case of the persona-related survey, which pertained to the trials with CloChat, we conducted a one-way repeated-measures ANOVA focusing on the types of situational contexts. To further explore significant findings, we applied Tukey’s

Table 2: List of situational scenarios used in our main study (Section 5). For each situation category (informational, emotional, and appraisal), we generated 1,000 scenarios and picked the representative ones using stratified sampling (Section 5.4).

Topic	Situation	Description
Informational support	Understanding company culture	<i>Users want to get the feel of a company's culture. A conversational agent helps by talking about the basics of corporate culture, what's unique to that company, and how to research more about it</i>
	Handling Stress	<i>Users look for ways to manage daily stress. A conversational agent shares techniques to relieve stress, advice on mental well-being, and other useful resources.</i>
	Exploring recipes	<i>Users keen on trying new dishes while chatting with a conversational agent about cooking methods, ingredients, and handy cooking tips</i>
	Travel planning	<i>Users plan trips by chatting with conversational about preparing for travel, cool places to visit, food spots to try, and useful local tips.</i>
Emotional support	Talking About Self-compassion	<i>When users are too hard on themselves, a conversational agent encourages them to be kinder to themselves and offers ways to practice self-esteem</i>
	Discussions on sleep issues	<i>Users having trouble sleeping want to talk with a conversational agent for understanding and suggestions on how to sleep better</i>
	Managing nightmares	<i>For users bothered by bad dreams, a conversational agent offers comfort and suggestions on managing them better.</i>
	Advice on romantic relationships	<i>Users facing romantic troubles talk with a conversational agent. In response, the agent offers understanding and tips for maintaining a healthy relationship.</i>
Appraisal support	Assessing my Skills and personal growth	<i>Users want to earn feedback on their academic or job skills by chatting with conversational agents. Users especially want to figure out strengths, areas to work on, and goals for personal growth.</i>
	Improving problem-solving skills	<i>Users discuss with a conversational agent how to think more logically and make better decisions.</i>
	Evaluating and building leadership skills	<i>Users who want to be better leaders discuss leadership styles, effective leadership practices, and ways to improve leadership with a conversational agent.</i>
	Boosting Project management skills	<i>Users chat with a conversational agent about how to manage projects better, from scheduling to working well with a team.</i>

Honestly Significant Difference (HSD) test [82] for post hoc analysis.

Results and Discussions. The results of our survey are depicted in Figure 7 (a-c), and a detailed result of statistical analysis is available in Appendix B. In the system-related survey (questions Q1–Q5, Q8–Q9, and Q11–Q15), we observed a significant main effect related to system types, as shown in Figure 7 (a). Our post hoc analysis indicated that CloChat consistently scored higher than ChatGPT across these questions. Although no significant main effects were detected for questions Q6 ('This system can provide useful answers to me.'), Q7 ('This system helps me achieve my goals.'), and Q10 ('I feel that this system will make my life more convenient.'), which focus on the perceived utility of conversational agents (Table 1), the trend still favored CloChat with higher average ratings.

These findings suggest that CloChat's personalized persona contributes positively to various user experience aspects, such as satisfaction, engagement, and future interaction likelihood. While statistically significant differences in perceived utility items were not observed, a consistent preference for CloChat was evident.

Regarding situation types in the system-related survey, significant main effects were noted for questions Q1, Q5–Q7, Q10–Q11, and Q14–Q15 (Figure 7 (b); detailed statistics reported in Appendix

B). Post hoc analysis showed that informational situations (Q5–Q7, Q10–Q11, and Q15) garnered higher scores compared to emotional situations, particularly in questions related to system effectiveness, utility, and future use intention. This indicates that users generally perceive conversational agents as more useful and effective for informational support than for emotional support, a trend independent of the presence of personalized personas.

In the persona-related survey, conducted exclusively with CloChat, significant main effects due to situation types were found in Q10 ('I can utilize this persona for work or academic purposes.') and Q12 ('It feels like this persona has a personality.') (Figure 7 (c); detailed statistics reported in Appendix B). The post hoc analysis of Q10 revealed significantly higher scores in informational situations than in emotional scenarios, aligning with the question's focus on the persona's utility in academic or professional settings. Conversely, in Q12, emotional situations scored higher than appraisal situations, emphasizing the human-like attributes and emotional resonance of the persona in these contexts.

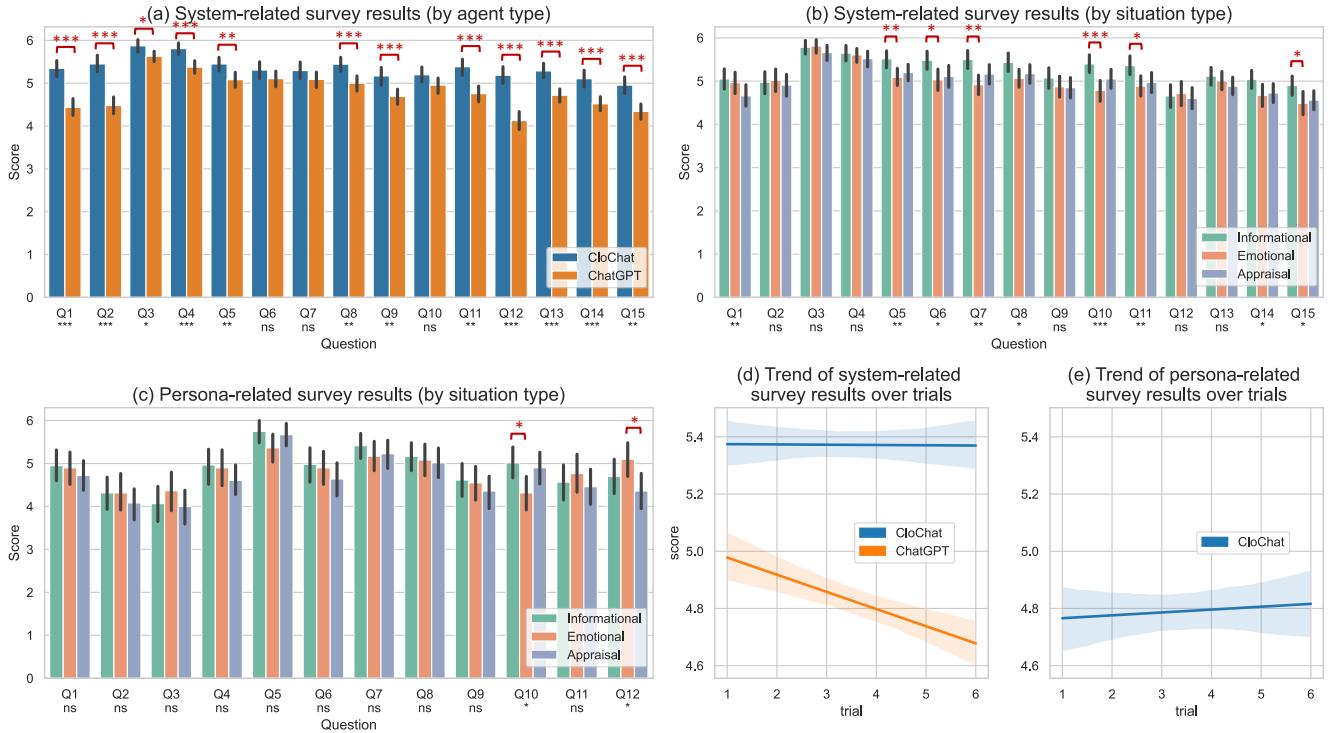


Figure 7: Post-trial survey results (Section 6.1, 6.2). **a, b)** Results of the system-related survey (Table 1a), aggregated by system type and situation type. **(c)** Results of the persona-related survey (Table 1b). **(d, e)** Trends in the system- and persona-related survey scores over trials. For the bar charts (a–c), the asterisks under each question number depict the statistical significance of the repeated-measures analysis of variance values (***: $p < .001$, **: $p < .01$, *: $p < .05$). Statistical significance found in the post hoc analysis is depicted with red brackets.

6.2 Temporal Evolution of Survey Scores Across Trials

Objectives. The objective of this analysis was to investigate the longitudinal changes in user evaluations of the conversational systems across multiple trials, addressing RQ1-2. Recognizing that higher scores in survey questions could be indicative of a better user experience, we sought to analyze the trends in overall user satisfaction over time.

Analysis design. To visually examine the temporal evolution of survey scores, we conducted regression analyses, plotting distinct regression lines for each conversational agent (CloChat and ChatGPT). For the system-related survey, we utilized an Analysis of Covariance (ANCOVA) [45] to statistically assess the significance of the observed differences in score trajectories between the two systems.

Results and Discussions. As depicted in Figure 7 (d), survey scores for ChatGPT showed a general downward trend over time, whereas scores for CloChat remained relatively stable. The ANCOVA analysis confirmed that the difference in these trends between CloChat and ChatGPT was statistically significant ($F = 89.89$; $p < .001$). Interestingly, the persona-related survey scores, gathered exclusively from CloChat trials, exhibited a slight upward trajectory. These findings suggest that while user experience with conversational

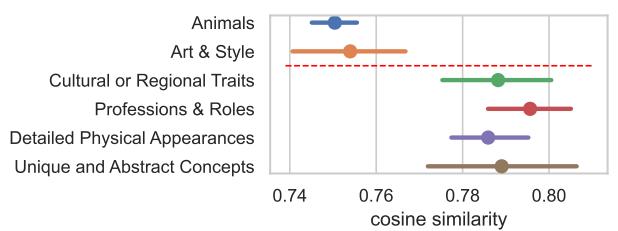


Figure 8: Degree of alignment between persona characteristics (non-visual traits) and visual traits based on the category of visual traits (Table 3). Higher cosine similarity scores represent better alignment. Post hoc analysis revealed that the categories on the upper side of the red dashed line obtained significantly lower cosine similarity scores than those on the lower side.

agents may typically decline over time, the presence of customized personas in CloChat appears to mitigate this effect, contributing to a sustained or even improved user experience.

Table 3: The categorization of visual traits discovered in our study. Our analysis (Section 6.3) shows that traits in *Animals* and *Art & Styles* categories tend to align less with the persona characteristic compared to the traits in the other categories.

Category	Sub-Category	Example codes	Count
<i>Animals</i>	Cute animals	Cute cat, Cute puppy, Cute bear, Cute panda, Cute seal	17
	Specific breeds	Korean Shorthair cat, Golden Retriever, Border Collie	3
	Animal behavior or moods	Wagging tail, Smile	3
<i>Cultural or Regional Traits</i>	Asian influences	Korean, East Asian woman, Vietnamese merchant	10
	Western influences	British gentleman, White Western woman	3
<i>Professions & Roles</i>	Business-related roles	Company employee, Executive in a startup, Office worker	6
	Academic professions	Professor, Graduate Student	5
	Service roles	Chef, Butler, Counselor, Doctor, Guide	7
	Creative roles	YouTuber, Actress, Musician	3
<i>Detailed Physical Appearances</i>	Hair types & styles	Short hair, Perm, Bald, Beard	9
	Clothing & accessories	Suit, Doctor's gown, Hat, Glasses, Pajamas, Hawaiian shirt	9
	Age	Baby, Middle-aged, in their 20s/30s/40s/50s	9
	Expressions and demeanor	Smiling, Serious, Tough, Friendly,	4
<i>Art & Style</i>	Painting styles	2D, 3D, Oil painting, Disney style	8
	Settings & background	Amusement park, Forest, Office	3
	Mood or emotion	Bright, Cute, Comfortable, Mysterious	4
<i>Unique and Abstract Concepts</i>	Quirky ideas	Cyber Buddha, Virgin Mary with electric guitar, Zhuangzi with wine	3
	Descriptive traits	Diligent, Hardworking, Charismatic, Kind	9
	Non-human characters	Rock, Blue square, Ghost	3

6.3 Alignment between the Visual and Non-visual Traits of Agent Personas

Objectives. In addressing RQ2 in a detailed way, we aimed to examine the correlation between the visual representations and non-visual characteristics (such as personality traits and roles) of customized agent personas. Existing literature suggests that a conversational agent's visual appearance often correlates with its persona, where specific traits or roles influence its visual depiction [59, 69, 88]. Our goal was to delve into this alignment, exploring how individuals intentionally coordinate these visual elements with their personalized agent personas.

Analysis design. We began with axial coding to categorize the relationship between various traits and the visual representations of agent personas. Two researchers independently created codebooks, which were then merged after discussions for consistent analysis.

To understand how the relationship between visual and non-visual traits varies across different visual trait categories, we first identified agent personas from our study where the visual representation fell into specific categories.

Next, we converted the visual and non-visual traits of these agent personas into vector embeddings. For visual traits, we used OpenAI's text-embedding API (with the text-embedding-ada-002 model) to transform image prompts into vectors. For personality traits, we transformed the natural language directives used in GPT-4 (refer to Section 4.3) into vector embeddings. We then calculated the cosine similarity between the vectors representing the image prompts and those representing the persona characteristics. A one-way ANOVA was conducted to assess differences in similarity scores across categories, followed by a post hoc analysis using Tukey's HSD test.

Results and discussions. Our analysis yielded six distinct categories of traits associated with visual representation: *Animals*, *Cultural or Regional Traits*, *Professions & Roles*, *Detailed Physical*

Appearances, *Art & Style*, and *Unique & Abstract Concepts* (for detailed coding results, see Table 3). A one-way ANOVA revealed significant differences in cosine similarity scores among these categories ($F(5, 148) = 8.190, p < .001$). Post hoc analysis using Tukey's HSD identified notably lower scores in the *Animals* and *Art & Style* categories compared to the others (Figure 8). For more detailed statistical information (p -values and confidence intervals), please see Appendix B.

A key distinction between categories with high and low similarity scores is the direct relevance of visual traits to human characteristics. Categories such as *Professions & Roles* (including specific roles like Office Worker, Professor, YouTuber) and *Cultural or Regional Traits* category (e.g., Korean, British) explicitly denote human subgroups, while the *Detailed Physical Appearances* category focuses on human features. Similarly, the *Unique and Abstract Concepts* category generally relates to human attributes, barring some non-human focused subcategories. In contrast, the *Art & Style* and *Animals* categories predominantly include traits that do not directly correspond to human attributes.

The results indicate that when participants chose visual traits closely linked to real-world human characteristics for their agent personas, there was a greater likelihood of alignment between these visual elements and the agent personas' non-visual traits. This tendency might also suggest that users often perceive their agent personas as virtual humans, expecting them to visually mirror typical human characteristics. Conversely, traits not directly related to human attributes tend to be applied more flexibly, reflecting individual user preferences rather than a strict alignment with the non-visual traits of their agent personas.

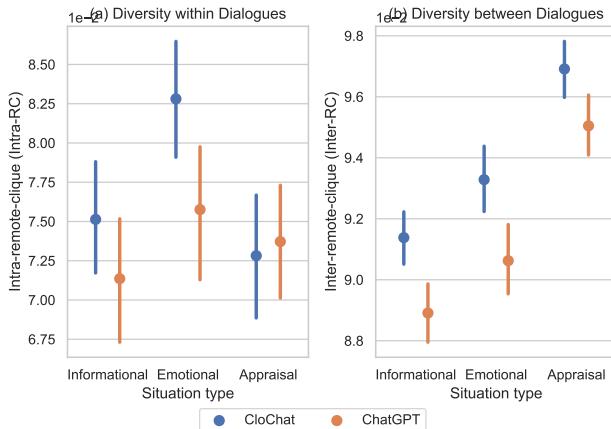


Figure 9: Diversity of dialogues created in our study, assessed by inter- and intra-remote-clique measures (Section 6.4). In summary, the dialogues with CloChat showed substantially higher diversity than those with ChatGPT.

6.4 Diversity of Dialogues

Objectives. In relation to RQ2, we hypothesized that using CloChat would lead to more enriched and diverse dialogues with conversational agents compared to standard ChatGPT interactions. Our goal was to empirically validate this hypothesis and explore the impact of different situational scenarios on the diversity of dialogues.

Analysis design. To rigorously evaluate dialogue diversity, we developed two specialized metrics: intra-remote-clique (intra-RC) and inter-remote-clique (inter-RC). These metrics are adaptations of the remote-clique (RC) metric [71], which is commonly used to measure text embedding diversity. The RC metric is defined as the average pairwise distance between text embeddings [26, 44]. Intra-RC specifically measures the average pairwise distance between utterances within a single dialogue, providing insight into the diversity of conversation within one session. Inter-RC, on the other hand, assesses the average linkage between utterances across two dialogues within the same situational context, offering a perspective on the diversity between different conversations under similar circumstances.

For each dialogue in our study, we computed the intra-RC to determine the level of diversity within that dialogue. We also calculated the inter-RC for each pair of dialogues sharing the same situational context to evaluate the diversity between conversations. To ensure that our metrics were not influenced by the semantic differences between various scenarios, we avoided comparing dialogues from distinct scenarios. We then conducted a two-way ANOVA to analyze the effects of system type (CloChat and ChatGPT) and situation type (informational, emotional, and appraisal support) on dialogue diversity. Tukey's HSD test was carried out for the post hoc analysis.

Results and discussions. The findings from our analysis are depicted in Figure 9. In terms of post hoc analysis, please refer to Appendix B. For intra-RC, a significant main effect was observed

for system types ($F(1, 354) = 4.16, p < .05$). However, post hoc analyses did not reveal any statistically significant differences between CloChat and ChatGPT. Regarding situation types, a significant main effect was also noted ($F(2, 354) = 6.13, p < .01$). Post hoc tests showed that dialogues in emotional contexts exhibited significantly higher diversity compared to both informational ($p < .01$) and appraisal ($p < .01$) contexts. We did not identify any interaction effects between system and situation types.

In the case of inter-RC, there were notable main effects for both system types ($F(1, 5214) = 30.91, p < .001$) and situation types ($F(2, 5214) = 67.71, p < .001$). Post hoc analysis revealed that dialogues using CloChat displayed a significantly higher level of diversity compared to ChatGPT ($p < .001$). Furthermore, we observed a systematic increase in dialogue diversity across the informational, emotional, and appraisal scenarios, with statistically significant differences in all pairwise comparisons ($p < .001$ for each). Again, no interaction effects were found.

To summarize, the results indicate that CloChat significantly enhanced the diversity of dialogues between different conversations (inter-dialogue diversity), but did not have a marked effect on the diversity within individual conversations (intra-dialogue diversity), in comparison to standard ChatGPT interactions. This suggests that while CloChat's tailored agent personas contribute to personalizing conversations, they may not necessarily increase the dynamic range of topics or conversational patterns within a single dialogue session.

7 QUALITATIVE RESULTS

In addition to our quantitative analysis, we delved into qualitative data from interviews to gain deeper insights into our research questions. We employed thematic analysis [15] as our methodological framework for the analysis. The research team utilized a line-by-line open coding technique, allowing for the identification and categorization of emergent themes from the interview data. The findings from this thematic analysis are detailed in the subsequent sections.

7.1 Patterns in Customizing and Selecting Agent Personas

Our user study revealed two distinct patterns in the creation and reuse of agent personas, each illustrating unique approaches to user engagement and satisfaction. The first pattern is characterized by dynamic persona customization, specifically tailored to meet immediate situational needs. Participants following this approach proactively envisioned specific scenarios for interaction and selected personas with appropriate characteristics, like personality and expertise, to match these situations. On average, participants in this group changed their agent personas 4.6 times over the six trials with CloChat, with more than half using six different personas for each session. For example, Participant 20 created a ‘psychiatrist’ persona to address stress and sleep concerns, commenting, “*I was super stressed, so I thought, why not talk to a ‘psychiatrist’?*” Similarly, in career guidance scenarios, participants customized personas to mimic employees from companies of interest, reflecting the importance of contextually relevant and personalized conversational experiences.

Conversely, the second pattern indicates a preference for reusing specific agent personas that have previously provided satisfactory conversational experiences. In our study, 12 participants consistently reused a particular persona for more than two trials, with some using the same persona throughout all six trials. For example, P11 repeatedly chose the ‘gentleman persona,’ stating, “*I kept using the ‘gentleman’ because he just gets me. He always knows the right thing to say.*” This pattern suggests that once a persona resonates with a user’s expectations, it fosters a sense of trust, reinforcing the user’s initial choice and encouraging future interactions. P27, for example, continued using a persona initially selected on a whim due to its unexpectedly accurate responses, saying, “*At first, I picked the persona just for kicks. But it was so on point, I kept coming back.*”

These two patterns differ fundamentally in their approach: the first is dynamic, with participants varying persona characteristics to suit specific needs, while the second is consistent, favoring a particular persona based on personal satisfaction and preference. This dichotomy illustrates how individual user preferences and needs can manifest in diverse ways when engaging with conversational systems, balancing between situational diversity and consistent personal preferences.

7.2 Conversation Diversity and Dynamics

The study revealed that the use of agent personas in conversational agents can offer a more diverse and enriched dialogue experience for participants. Initially, some participants expressed during interviews that they primarily utilized LLMs for basic tasks like answering simple questions or conducting fundamental information searches, valuing ChatGPT’s immediate response capabilities over complex customization options. However, post-experiment interviews revealed a notable shift in perception. P8 observed, “*Even if the answers are the same, having a persona adds a more professional feel. I think it could be useful even in casual conversations.*” The comment suggests that customized agent personas can influence their user experience in a positive way, indicating a potential shift in user behavior from basic information retrieval to seeking more personalized and engaging interactions.

Despite the experiment’s scenarios being categorized as informational, emotional, and appraisal, participants often ventured beyond these confines. Their intrigue with personalized agent personas led them to explore new topics and questions. P13 reflected, “*The conversation got longer when I found more fun and interesting topics, similar to talking with friends. With ChatGPT, the conversations were shorter due to predictable responses.*” This expansion in dialogue scope fostered deeper and more intricate relationships between participants and agent personas.

The agent personas not only influenced the nature of the dialogue but also affected the participants’ conversational styles. Engaging with a friendly and humorous persona, for example, fostered a light-hearted atmosphere, encouraging participants to use informal language and share jokes. P27 noted, “*Talking to this persona felt like chatting with an old friend. I often found myself laughing.*” Conversely, interactions with more serious or formal personas led to dialogues with a scholarly or cautious tone. P13 commented, “*My persona was cold and academic, like Sherlock Holmes, which naturally steered the conversation to be more serious.*” These dynamics even

impacted the participants’ moods and emotions, as highlighted by P30: “*I felt more energetic talking to my vibrant persona, whereas serious conversations prompted deeper thought.*”

7.3 Relationship between Participants and Agent Personas

The introduction of agent personas led participants to perceive their conversational partners as entities with unique contexts and personalities, rather than just as programs. Many participants reported enhanced immersion and trust in their interactions when the agent persona’s responses aligned with their expectations or preferences. For example, P1 expressed, “*Having a personalized persona made the conversation feel more alive, and I felt more trust in the interaction.*” This increase in trust, as evidenced by a previous study [49], highlights the importance of persona alignment in fostering meaningful conversational experiences.

In contrast, interactions with ChatGPT were often perceived as engaging with an automated responder, lacking a personal touch. Participants like P13 remarked, “*My conversations with ChatGPT felt pretty standard. It was like getting necessary information from a machine, without any specific expectation or connection.*” This difference underscores the uniqueness and personalization that agent personas can bring to conversational experiences.

Another significant aspect of our findings pertains to the emotional connection participants developed with their configured personas. Some participants experienced profound emotional responses during their interactions. P6 shared, “*The conversation moved me almost to tears,*” while P19 described the conversation as akin to talking with a friend due to the persona’s empathy.

The visual representation of personas also played a crucial role in enhancing empathy and engagement. P15 mentioned, “*Seeing the persona I created made the conversation feel more direct, eliciting a deeper sense of empathy.*” The act of visualizing and personalizing these personas enriched the conversational experience, as P9’s comment illustrates: “*I crafted it thinking of my favorite YouTuber. During our chat, I imagined his voice, making the conversation more engaging.*” This aligns with research findings that emphasize the power of visual engagement in enhancing conversational interest [83].

Initially, many participants were not inclined to use conversational agents for emotional support, a trend also supported by our quantitative findings (Section 6.1). However, as the experiment progressed, participants began to appreciate the value of emotional conversations with agent personas. P19’s reflection captures this shift: “*The experiment taught me the value of emotional conversations with conversational agents. CloChat’s personalized agents responded warmly, understanding my feelings remarkably well.*”

7.4 User Feedback on the Persona Customization

The participants found that CloChat’s form-based interface significantly lowered the entry barrier for engaging with conversational agents, making it more accessible to the general public. During pre-experiment interviews, many participants revealed difficulties due to limited technical knowledge needed for LLM customization, particularly when it came to selecting specific characteristics

for personas. Thus, for participants unfamiliar with crafting text prompts, the availability of predefined persona trait options in CloChat was notably more user-friendly. P17, who had initially been concerned about the complexity of prompt creation, observed after the experiment, “*CloChat definitely reduces the effort needed to create a persona. It’s convenient not having to think about specific text prompts.*” As the trials progressed, participants developed their own strategies for effectively customizing unique personas. P30 commented, “*Customizing personas was initially challenging, but I quickly discovered the optimal approach.*” This feedback indicates that users experienced a manageable learning curve with the CloChat interface.

Nevertheless, some participants pointed out that setting up personas could be complicated and time-consuming without clear guidelines or presets. P15 noted, “*I was a bit confused when first setting up the persona. I wasn’t sure how to approach it or what criteria to use for selection.*” While most acknowledged the benefits of having bespoke personas, there were mentions of the burden involved in their initial setup as well, suggesting a need for more user-friendly guidance or preset options.

The feature allowing users to customize visual representations of agent personas was particularly appreciated, offering an enhancement not found in ChatGPT. P13 remarked, “*Modifying the counselor’s appearance was surprising and greatly enhanced my engagement.*” This emphasizes the vital role of visual representation in the design and functionality of conversational agents, enhancing user engagement and expectation management.

7.5 Reflecting Real Life to Agent Personas

A notable trend among participants was the incorporation of elements from their real-life experiences and observations into their agent personas, rather than creating entirely fictional characters. For instance, participants often modeled personas after familiar individuals like acquaintances, friends, pets, or celebrities. P9, who chose a renowned doctor as a persona, shared, “*I based the persona on a real person I saw on TV. Reflecting his tone in my agent persona made the conversation warmer and more immersive, allowing me to speak more honestly.*” Similarly, P27 created a persona inspired by a friend’s occupation and hobbies, noting, “*Seeing these characteristics in the conversation gave it the feeling of talking to my actual friend.*” This approach illustrates how personal experiences can enhance the realism and relatability of conversational partners. However, this practice can also raise ethical concerns regarding privacy and personal data protection, as it involves imitating or mimicking real individuals potentially without their consent.

Participants also enjoyed the imaginative exercise of setting up their pets as personas, attributing them with imagined personality traits and habits. P2 reflected, “*I mirrored my dog’s playful personality. Imagining his responses made the conversation more fun and unique.*”

The practice of drawing from real-life experiences for persona customization allowed participants to infuse their personal lives and emotional connections into the digital domain. Nevertheless, while this approach significantly enriches user interaction with

conversational AI systems, it simultaneously highlights the importance of addressing ethical considerations related to mimicking real-world individuals.

8 DISCUSSIONS

Our user study was aimed at investigating the impact of agent persona customization on user experience during interactions with LLM-based conversational agents, as opposed to conventional generic conversational agents (RQ1). We discovered that the customization of agent personas significantly boosts user engagement, trust, and emotional connection, offering a noticeable improvement in maintaining user satisfaction and engagement compared to ChatGPT. In addressing RQ2, we delved into the ways users customize their agent personas and the resultant effects on their interactions. We observed that conversations involving customized agent personas tend to be richer and more diverse. Users often align the traits of agent personas in terms of both visual elements and real-world inspirations, which additionally brings to light ethical considerations regarding agent persona customization. In extending our discussions on these findings, we explore relevant topics and present practical implications for the design of user interfaces employing LLM-based conversational agents. We also outline the limitations of our study, acknowledging areas that could benefit from further exploration and improvement.

8.1 The Multifaceted Roles of Customizable Agent Personas

Our study demonstrated that CloChat provided an enhanced user experience compared to ChatGPT, highlighting the substantial benefits and potential of customizable agent personas. Users interacting with CloChat perceived the agent personas not just as algorithmic tools, but as distinct conversational partners with unique personalities, as outlined in (Section 7.3). This shift in perception, supported by previous research [48, 49, 56], increased users’ emotional engagement, trust, and immersion in the conversational experience.

A noteworthy observation was how some participants modified their own conversational styles to resonate more with the personas they created, indicating a deepening emotional connection with their customized agents (Section 7.3). The integration of visual representations further solidified this bond, elevating the agents from mere information retrieval tools to authentic conversational partners (Section 7.3, Section 6.3).

Conversely, interactions with ChatGPT were associated with lower levels of emotional engagement (Section 7.3). This contrast not only underscores the limitations of text-prompt-focused platforms like ChatGPT but also highlights the potential of CloChat’s comprehensive personalization features. These features can have the ability to enrich user experiences across diverse emotional contexts and situations.

In conclusion, the customizable agent personas in CloChat extend beyond traditional information retrieval roles typically associated with conversational agents using LLMs. They play a crucial role in fostering emotional connections and enhancing user engagement with conversational systems, indicating an expansion in both the functional scope and emotional depth of these technologies.

8.2 Personas' Role in Sustaining User Engagement on Conversational Agents

While ChatGPT is renowned for its conversational capabilities, it faces limitations in reflecting users' individual preferences and sustaining deep, ongoing relationships, as it primarily excels in basic information retrieval and short interactions [13]. Our study confirms this, indicating a decline in user satisfaction with ChatGPT over time (Section 6.2).

In contrast, personalized agent personas not only elicited initial positive responses from users but also played a pivotal role in maintaining these positive connections over time (Section 6.2). This aligns with prior research [14, 85] and our qualitative findings (Section 7.1), suggesting that user preferences are dynamic, varying according to mood, situation, and context [81]. CloChat's capability to customize a variety of personas to adapt to these shifting preferences likely contributed to sustained user engagement.

Another key factor in the enduring positive relationship with personalized agent personas is the human-like perception they create (Section 7.3), resonating with findings from Cowan et al. [21]. With CloChat, participants engaged in longer conversations and explored a wider range of topics (Section 7.2, 6.4), leading to increased trust and satisfaction. This enriched conversational experience contributes to sustainable interaction with the agent, moving beyond brief, transactional conversations. Although our study did not specifically observe long-term interactions between users and customized agents, the implications from our findings hint at the potential for fostering lasting relationships with conversational agents in the future.

8.3 Pros and Cons of Persona Customization

Our study underscores the significant advantages of incorporating persona customization features into LLM-based conversational user interfaces. The majority of participants responded positively to this functionality, noting that it made their conversations more enjoyable and engaging (Section 7.3, Section 7.4). The ability to tailor personas according to personal preferences fostered increased interest and active participation in conversations, leading to a more open and dynamic interaction, as reflected in survey responses (Section 6.1).

However, alongside these benefits, certain challenges were also observed (Section 7.4). Some participants found the wide array of customization options to be overwhelming, particularly for those new to conversational agents or not versed in prompt engineering techniques. To address this, future iterations could consider integrating automated suggestions that assist users in managing their expectations and simplifying the decision-making process. This could involve methods like OpenAI's recently released GPTs, which can learn specific knowledge or personalities from user-provided documents [6]. Further research is needed to compare various approaches, such as extensive user-driven customization versus agents automatically learning from user documents, and to understand how these different methods influence user experience. An effective balance between user-driven customization and automated recommendations, as suggested in literature [46], could provide a solution to these challenges.

The overarching aim would be to streamline the customization process, making it less daunting for users while still offering a rich, personalized experience. This balance is key to harnessing the full potential of persona customization in enhancing user engagement with conversational AI systems.

8.4 Ethical Concerns on Personalized Personas

In our study, we observed that participants frequently drew inspiration from their personal experiences and daily interactions when customizing their agent personas (see Section 7.5). A notable trend involved mimicking celebrities or personal acquaintances. This inclination could be attributed to the perceived expertise or symbolic stature of famous individuals or a preference for replicating interactions with familiar and relatable figures rather than inventing entirely new or unknown personas. While this method can lend a sense of realism to interactions with conversational agents and potentially foster more robust and lasting connections, it also brings forth significant ethical dilemmas.

This practice might risk privacy breaches and confidentiality issues, particularly when integrating distinct details or characteristics of these individuals, such as their occupation, location and relationships with others. Furthermore, since a persona cannot fully encompass the complexity of an actual person's personality, actions, or thoughts, such representations may lead to misconceptions or biases. These misrepresentations could adversely impact the reputations or identities of the individuals portrayed, as discussed in the research by Deshpande et al. [8]. Hence, a delicate balance must be struck between the creative liberty in persona customization and the ethical implications of drawing from real-life figures.

In the context of LLMs operating across networks, using personal information to shape agent personas raises concerns about individual privacy. Once personal identifying data is input into an LLM, its permanence and the opaque nature of data storage and processing can result in unintended privacy violations, with interactions potentially reaching a broad, unknown audience.

Echoing the observations of Goldstein et al. [36], the realm of AI ethics is continuously evolving. Ongoing dialogue and development are essential to establish ethical frameworks and principles within this field. Consequently, it's critical to develop practical and robust solutions for ethical issues related to language model applications. Researchers and developers should diligently address these ethical aspects in the design and deployment of personas, implementing safeguards to protect personal information during the training of machine learning models. This step is fundamental to preserving user privacy and ensuring the ethical use of LLM-based conversational systems. Moreover, clear ethical guidelines and protocols for persona design are necessary. Users should be informed about the risks of imitating real individuals and discouraged from engaging in such practices. Future research should delve into the potential problems of using personas based on real individuals in specific contexts. It may be advisable to limit the use of personas based on real people, especially in scenarios requiring expert advice or sensitive discussions (e.g., sexual dialogue). Such measures will help users grasp the ethical implications of their choices and encourage responsible persona creation.

8.5 Design Implications

Based on our discussions, we propose the following design implications for future development and refinement of conversational user interfaces employing LLMs:

- **I1: Prioritize Persona Customization to Enhance User Trust and Engagement.** CloChat surpassed ChatGPT in terms of user satisfaction, largely owing to the availability of customizable personas. Designers should, therefore, consider prioritizing persona customization options in their systems. The heightened user trust and improved conversation quality associated with personalized personas highlight their vital role in the future design of conversational interfaces.
- **I2: Minimize the Initial Setup Burden to Encourage User Engagement.** The initial setup for persona customization can be perceived as burdensome (Section 8.3). Designers should streamline the setup process and provide easy-to-follow onboarding assistance, thus enhancing user immersion and engagement from the outset (Section 8.1, 8.2). This could involve introducing conversational tutorials or preset persona options.
- **I3: Make the Agent Persona Adaptive.** The flexibility in creating bespoke agent personas for various situations could be helpful for sustaining user engagement with CloChat. We recommend implementing adaptive algorithms that tailor persona behaviors based on user intentions and circumstances, combined with easy customization options. This approach would cater to users who prefer consistent personas across different scenarios as well as those who desire situation-specific persona adaptations.
- **I4: Provide Thorough Guidelines on Ethical Considerations.** The study revealed potential ethical issues, such as incorporating celebrities or real-life acquaintances into agent persona designs without consent. Given the likelihood of further ethical concerns, it is crucial to provide users with clear guidelines addressing these issues. This will help ensure that the customization of agent personas adheres to ethical standards and respects individual privacy and rights.

8.6 Limitations and Future Work

Our study, while shedding light on the diverse user experiences with customizable agent personas in LLMs, has several limitations that must be acknowledged. Firstly, the participant pool was limited to Korean speakers, due to institutional constraints. This limitation may affect the generalizability of our findings to other linguistic and cultural groups. Future studies should aim to include a more diverse range of participants to broaden the applicability of the results. Secondly, the creation of agent personas in CloChat relied solely on prompt injection, which may lack depth in specialized or rapidly evolving domains [5]. This limitation raises the question of how LLMs can be optimized for more in-depth and accurate persona representations. Future research could explore advanced techniques such as fine-tuning [7] or the integration of external memory [66, 76] to enhance the sophistication of persona customization in LLMs. Thirdly, CloChat's persona customization is currently confined to a form-based interface. While this design choice was made to lower the barrier to persona customization,

it is worth exploring how different customization methods (e.g., direct prompt writing, conversation-based customization [4]) might impact user experience. These alternative approaches could offer more flexibility and personalization, catering to users with varying levels of expertise and preferences. Lastly, our study did not explore the long-term user experience with CloChat. To fill this gap, future research endeavors should focus on longitudinal studies to understand how user engagement with CloChat evolves over time. Such studies are crucial for uncovering the distinctions between short-term and long-term interactions and for developing strategies to cultivate sustained and meaningful relationships with conversational agents like CloChat.

9 CONCLUSION

In this study, we explored how users engage with and customize agent personas in LLM-based user interfaces. For this purpose, we developed CloChat, a user-centric interface built upon ChatGPT, enabling users to easily customize and interact with agent personas. We then compared CloChat with the standard ChatGPT system through a user study. Our findings indicate that CloChat significantly improves the overall user experience compared to ChatGPT, suggesting that giving users the ability to personalize their conversational agents leads to a more satisfying experience. Additionally, it was observed that users not only enjoy the process of customizing their agents but also find these personalized agents to be more engaging conversational partners. Users developed more meaningful relationships with the customized personas and engaged in more prolonged interactions with them. Drawing from these insights, we proposed design implications for future systems utilizing conversational agents. We hope this will pave the way for groundbreaking advancements in the design of conversational interactions facilitated by LLMs.

ACKNOWLEDGMENTS

This research was supported by the Yonsei University Research Fund of 2023 (2023-22-0430) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2023R1A2C200520911). This work was also supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)] The ICT at Seoul National University provided research facilities for this study.

REFERENCES

- [1] [n. d.]. F/awesome-CHATGPT-prompts. <https://github.com/f/awesome-chatgpt-prompts>
- [2] [n. d.]. Google Bard. <https://bard.google.com/>
- [3] [n. d.]. Introducing ChatGPT. <https://openai.com/blog/chatgpt/>
- [4] [n. d.]. Introducing GPTs. <https://openai.com/blog/introducing-gpts>
- [5] Ankush Agarwal, Sakharam Gawade, Amar Prakash Azad, and Pushpak Bhattacharyya. 2023. KITLM: Domain-Specific Knowledge InTegration into Language Models for Question Answering. arXiv:2308.03638 [cs.CL]
- [6] Open AI. 2023. *Introducing GPTs*. <https://openai.com/blog/introducing-gpts>
- [7] Nikolicich Alexandr, Osiakova Irina, Kudinova Tatyana, Kappusheva Inessa, and Puchkova Arina. 2021. Fine-Tuning GPT-3 for Russian Text Summarization. In *Data Science and Intelligent Systems*, Radek Silhavy, Petr Silhavy, and Zdenek Prokopova (Eds.). Springer International Publishing, Cham, 748–757.
- [8] Deshpande Ameet, Murahari Vishvak, Rajpurohit Tanmay, Kalyan Ashwin, and Narasimhan Karthik. 2023. Toxicity in chatGPT: Analyzing persona-assigned

- language models. *arXiv preprint arXiv: 2304.05335* (2023). <https://doi.org/10.48550/arXiv.2304.05335>
- [9] Farshid Anvari, Deborah Richards, Michael Hitchens, Muhammad Ali Babar, Hien Minh Thi Tran, and Peter Busch. 2017. An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software* 134 (2017), 324–339. <https://doi.org/10.1016/j.jss.2017.09.020>
- [10] Farshid Anvari, Deborah Richards, Michael Hitchens, Muhammad Ali Babar, Hien Minh Thi Tran, and Peter Busch. 2017. An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software* 134 (2017), 324–339. <https://doi.org/10.1016/j.jss.2017.09.020>
- [11] Y Bang, S Cahyawijaya, N Lee, W Dai, D Su, B Wilie, H Lovenia, Z Ji, T Yu, W Chung, et al. 2023. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *arXiv*. <https://doi.org/10.48550/arXiv.2302.04023>
- [12] Simona Borsci, Alessio Malizia, Martin Schmettow, Frank Van Der Velde, Gunay Tariverdiyeva, Divyaa Balaji, and Alan Chamberlain. 2022. The Chatbot Usability Scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing* 26 (2022), 95–119. <https://doi.org/10.1007/s00779-021-01582-9>
- [13] Petter Bae Brandzaeg and Asbjørn Følstad. 2018. Chatbots: changing user needs and motivations. *interactions* 25, 5 (2018), 38–43. <https://doi.org/10.1145/3236669>
- [14] Michael Braun, Anna Mainz, Ronne Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11. <https://doi.org/10.1145/3290605.3300270>
- [15] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [16] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712* (2023). <https://doi.org/10.48550/arXiv.2303.12712>
- [17] Yen-ning Chang, Youn-kyung Lim, and Erik Stolterman. 2008. Personas: From Theory to Practices (*NordiCHI '08*). Association for Computing Machinery, New York, NY, USA, 439–442. <https://doi.org/10.1145/1463160.1463214>
- [18] Yen-ning Chang, Youn-kyung Lim, and Erik Stolterman. 2008. Personas: from theory to practices. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. 439–442. <https://doi.org/10.1145/1463160.1463214>
- [19] Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *International Journal of Human–Computer Interaction* 37, 8 (2021), 729–758. <https://doi.org/10.1080/10447318.2020.1841438>
- [20] Ssu Chiu, Maolin Li, Yen-Ting Lin, and Yun-Nung Chen. 2022. Salesbot: Transitioning from chit-chat to task-oriented dialogues. *arXiv preprint arXiv:2204.10591* (2022). <https://doi.org/10.48550/arXiv.2204.10591>
- [21] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. "What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services*. 1–12. <https://doi.org/10.1145/3098279.3098539>
- [22] Carolyn E Cutrona and Julie A Suh. 1992. Controllability of stressful events and satisfaction with spouse support behaviors. *Communication research* 19, 2 (1992), 154–174. <https://doi.org/10.1177/009365092019002002>
- [23] Hayco de Haan, Joop Snijders, Christof van Nimwegen, and Robbert Jan Beun. 2018. Chatbot personality and customer satisfaction. *Info Support Research* (2018). <https://research.infosupport.com/wp-content/uploads/Chatbot-Personality-and-Customer-Satisfaction-Bachelor-Thesis-Information-Sciences-Hayco-de-Haan.pdf>
- [24] Ameet Deshpande, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. 2023. Toxicity in chatgpt: Analyzing persona-assigned language models. *arXiv preprint arXiv:2304.05335* (2023).
- [25] Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, and Jason Weston. 2018. Wizard of wikipedia: Knowledge-powered conversational agents. *arXiv preprint arXiv:1811.01241* (2018). <https://doi.org/10.48550/arXiv.1811.01241>
- [26] Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2011. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy. *ACM Trans. Comput.-Hum. Interact.* 17, 4, Article 18 (dec 2011), 24 pages. <https://doi.org/10.1145/1879831.1879836>
- [27] Robin IM Dunbar, Anna Marriott, and Neil DC Duncan. 1997. Human conversational behavior. *Human nature* 8 (1997), 231–246. <https://doi.org/10.1007/BF02912493>
- [28] Nature Editorial. 2023. *ChatGPT is a black box: how AI research can break it open*. <https://www.nature.com/articles/d41586-023-02366-2>
- [29] Daniel Fernau, Stefan Hillmann, Nils Feldhus, and Tim Polzehl. 2022. Towards automated dialog personalization using mbti personality indicators. In *Proc. Interspeech*. 1968–1972. <https://doi.org/10.21437/Interspeech.2022-376>
- [30] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering cognitive behavioral therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health* 4, 2 (2017), e7785. <https://doi.org/10.2196/mental.7785>
- [31] Tabor E Flickinger, Claire DeBolt, Ava Lena Waldman, George Reynolds, Wendy F Cohn, Mary Catherine Beach, Karen Ingersoll, and Rebecca Dillingham. 2017. Social support in a virtual community: analysis of a clinic-affiliated online support group for persons living with HIV/AIDS. *AIDS and Behavior* 21 (2017), 3087–3099. <https://doi.org/10.1007/s10461-016-1587-3>
- [32] Luciano Floridi and Massimo Chiratti. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines* 30 (2020), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- [33] Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural approaches to conversational AI. In *The 41st international ACM SIGIR conference on research & development in information retrieval*. 1371–1374. <https://doi.org/10.1145/3209978.3210183>
- [34] Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. 2018. A knowledge-grounded neural conversation model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. <https://doi.org/10.1609/aaai.v32i1.11977>
- [35] Anand Goluk. 2023. LLMs and AI: Understanding Its Reach and Impact. (2023).
- [36] Josh A Goldstein, Girish Sastry, Micah Musser, Renee DiResta, Matthew Gentzel, and Katerina Sedova. 2023. Generative language models and automated influence operations: Emerging threats and potential mitigations. *arXiv preprint arXiv:2301.04246* (2023). <https://doi.org/10.48550/arXiv.2301.04246>
- [37] Isabel Kathleen Fornell Haugeland, Asbjørn Følstad, Cameron Taylor, and Cato Alexander Bjørkli. 2022. Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human-Computer Studies* 161 (2022), 102788. <https://doi.org/10.1016/j.ijhcs.2022.102788>
- [38] hu.ma.ne. 2023. *Ai Pin Overview*. <https://hu.ma.ne/aipin>
- [39] Youjin Hwang, Seokwoo Song, Donghoon Shin, and Joonhwan Lee. 2021. Linguistic Features to Consider When Applying Persona of the Real Person to the Text-Based Agent (*MobileHCI '20*). Association for Computing Machinery, New York, NY, USA, Article 23, 4 pages. <https://doi.org/10.1145/3406324.3410723>
- [40] Hyeon Jeon, Hyung-Kwon Ko, Jaemin Jo, Youngtaek Kim, and Jinwook Seo. 2022. Measuring and Explaining the Inter-Cluster Reliability of Multidimensional Projections. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 551–561. <https://doi.org/10.1109/TVCG.2021.3114833>
- [41] Hyeon Jeon, Hyung-Kwon Ko, Soohyun Lee, Jaemin Jo, and Jinwook Seo. 2022. Uniform Manifold Approximation with Two-phase Optimization. In *2022 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 80–84. <https://doi.org/10.1109/VIS54862.2022.000025>
- [42] Hyeon Jeon, Ghulam Jilani Quadri, Hyunwook Lee, Paul Rosen, Danielle Albers Szafir, and Jinwook Seo. 2023. CLAMS: A Cluster Ambiguity Measure for Estimating Perceptual Variability in Visual Clustering. *arXiv preprint arXiv:2308.00284* (2023). <https://doi.org/10.48550/arXiv.2308.00284>
- [43] Hang Jiang, Xiajie Zhang, Xubo Cao, Jad Kabbara, and Deb Roy. 2023. Personallm: Investigating the ability of gpt-3.5 to express personality traits and gender differences. *arXiv preprint arXiv:2305.02547* (2023). <https://doi.org/10.48550/arXiv.2305.02547>
- [44] Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Trans. Interact. Intell. Syst.* 7, 1, Article 2 (dec 2016), 42 pages. <https://doi.org/10.1145/2926720>
- [45] H. J. Keselman, Carl J. Huberty, Lisa M. Lix, Stephen Olejnik, Robert A. Cribbie, Barbara Donahue, Rhonda K. Kowalchuk, Laureen L. Lowman, Martha D. Petoskey, Joanne C. Keselman, and Joel R. Levin. 1998. Statistical Practices of Educational Researchers: An Analysis of their ANOVA, MANOVA, and ANCOVA Analyses. *Review of Educational Research* 68, 3 (1998), 350–386. <https://doi.org/10.3102/00346543068003350>
- [46] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22 (2012), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- [47] A Bakti Kocaballi. 2023. Conversational ai-powered design: Chatgpt as designer, user, and product. *arXiv preprint arXiv:2302.07406* (2023). <https://doi.org/10.48550/arXiv.2302.07406>
- [48] Sunok Lee, Sungbae Kim, and Sangsu Lee. 2019. "What does your Agent look like?" A Drawing Study to Understand Users' Perceived Persona of Conversational Agent. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6. <https://doi.org/10.1145/3290607.3312796>
- [49] Nadine Lessio and Alexis Morris. 2020. Toward Design Archetypes for Conversational Agent Personality. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3221–3228. <https://doi.org/10.1109/SMC42975.2020.9283254>

- [50] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055* (2015). <https://doi.org/10.48550/arXiv.1510.03055>
- [51] Jungwoo Lim, Myunghoon Kang, Yuna Hur, Seungwon Jung, Jinsung Kim, Yoonna Jang, Dongyub Lee, Hyesung Ji, Donghoon Shin, Seungryong Kim, et al. 2023. You Truly Understand What I Need: Intellectual and Friendly Dialogue Agents Grounding Knowledge and Persona. *arXiv preprint arXiv:2301.02401* (2023). <https://doi.org/10.48550/arXiv.2301.02401>
- [52] Li Liu and Vincent G Duffy. 2023. Exploring the Future Development of Artificial Intelligence (AI) Applications in Chatbots: A Bibliometric Analysis. *International Journal of Social Robotics* 15, 5 (2023), 703–716. <https://doi.org/10.1007/s12369-022-00956-0>
- [53] Shan Liu, Muyu Zhang, Baojun Gao, and Guoyin Jiang. 2020. Physician voice characteristics and patient satisfaction in online health consultation. *Information & Management* 57, 5 (2020), 103233. <https://doi.org/10.1016/j.im.2019.103233>
- [54] Pierre-Emmanuel Mazaré, Samuel Humeau, Martin Raison, and Antoine Bordes. 2018. Training millions of personalized dialogue agents. *arXiv preprint arXiv:1809.01984* (2018). <https://doi.org/10.48550/arXiv.1809.01984>
- [55] Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of personality* 60, 2 (1992), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- [56] Sara Moussawi and Raquel Benbunan-Fich. 2021. The effect of voice and humour on users' perceptions of personal intelligent agents. *Behaviour & Information Technology* 40, 15 (2021), 1603–1626. <https://doi.org/10.1080/0144929X.2020.1772368>
- [57] Sara Moussawi, Marios Koufaris, and Raquel Benbunan-Fich. 2021. How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets* 31 (2021), 343–364. <https://doi.org/10.1007/s12525-020-00411-w>
- [58] Tatwadarshi P Nagarhalli, Vinod Vaze, and NK Rana. 2020. A review of current trends in the development of chatbot systems. In *2020 6th International conference on advanced computing and communication systems (ICACCS)*. IEEE, 706–710. <https://doi.org/10.1109/ICACCS48705.2020.9074420>
- [59] Ha Nguyen. 2022. Examining Teenagers' Perceptions of Conversational Agents in Learning Settings. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (Braga, Portugal) (IDC '22). Association for Computing Machinery, New York, NY, USA, 374–381. <https://doi.org/10.1145/3501712.3529740>
- [60] S Nithuna and CA Laseena. 2020. Review on implementation techniques of chatbot. In *2020 International Conference on Communication and Signal Processing (ICCP)*. IEEE, 0157–0161. <https://doi.org/10.1109/ICCP48568.2020.9182168>
- [61] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375* (2023). <https://doi.org/10.48550/arXiv.2303.13375>
- [62] Inc Open AI. 2023. *Introducing chatgpt*. <https://openai.com/blog/chatgpt/>
- [63] OpenAI. 2023. GPT-4 Technical Report. <https://doi.org/10.48550/arXiv.2303.08774> [cs,CL]
- [64] Keyu Pan and Yawen Zeng. 2023. Do LLMs Possess a Personality? Making the MBTI Test an Amazing Evaluation for Large Language Models. *arXiv preprint arXiv:2307.16180* (2023). <https://doi.org/10.48550/arXiv.2307.16180>
- [65] Anshul Vikram Pandey, Josua Krause, Cristian Felix, Jeremy Boy, and Enrico Bertini. 2016. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3659–3669. <https://doi.org/10.1145/2858036.2858155>
- [66] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023). <https://doi.org/10.48550/arXiv.2304.03442>
- [67] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents (CUI '21). Association for Computing Machinery, New York, NY, USA, Article 12, 4 pages. <https://doi.org/10.1145/3469595.3469607>
- [68] Alisha Pradhan and Amanda Lazar. 2021. Hey Google, do you have a personality? Designing personality and personas for conversational agents. In *Proceedings of the 3rd Conference on Conversational User Interfaces*. 1–4. <https://doi.org/10.1145/3469595.3469607>
- [69] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. " Alexa is my new BFF" social roles, user satisfaction, and personification of the Amazon Echo. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 2853–2859. <https://doi.org/10.1145/3027063.3053246>
- [70] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3. <https://doi.org/10.48550/arXiv.2204.06125>
- [71] Samuel Rhys Cox, Yunlong Wang, Ashraf Abdul, Christian von der Weth, and Brian Y. Lim. 2021. Directed Diversity: Leveraging Language Embedding Distances for Collective Creativity in Crowd Ideation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 393, 35 pages. <https://doi.org/10.1145/3411764.3445782>
- [72] Mustafa Saifari, Greg Serapio-García, Clément Crepy, Stephen Fitz, Peter Romero, Luning Sun, Marwa Abdulhai, Aleksandra Faust, and Maja Matarić. 2023. Personality traits in large language models. *arXiv preprint arXiv:2307.00184* (2023). <https://doi.org/10.48550/arXiv.2307.00184>
- [73] Joni Salminen, Soon-gyo Jung, João M. Santos, Shammur Chowdhury, and Bernard J. Jansen. 2020. The Effect of Experience on Persona Perceptions. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3334480.3382786>
- [74] Joni Salminen, Haewoon Kwak, João M. Santos, Soon-Gyo Jung, Jisun An, and Bernard J. Jansen. 2018. Persona Perception Scale: Developing and Validating an Instrument for Human-Like Representations of Data. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI EA '18). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188461>
- [75] Joni Salminen, Joao M. Santos, Soon-Gyo Jung, Motahareh Esfandi, and Bernard J. Jansen. 2020. Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. *International Journal of Human-Computer Interaction* 36, 8 (2020), 788–800. <https://doi.org/10.1080/10447318.2019.1688946>
- [76] Dale Schuurmans. 2023. Memory Augmented Large Language Models are Computationally Universal. *arXiv:2301.04589* [cs,CL]
- [77] Emily Sheng, Josh Arnold, Zhou Yu, Kai-Wei Chang, and Nanyun Peng. 2021. Revealing persona biases in dialogue systems. *arXiv preprint arXiv:2104.08728* (2021). <https://doi.org/10.48550/arXiv.2104.08728>
- [78] Emily Sheng, Kai-Wei Chang, Premkumar Natarajan, and Nanyun Peng. 2020. Towards controllable biases in language generation. *arXiv preprint arXiv:2005.00268* (2020). <https://doi.org/10.48550/arXiv.2005.00268>
- [79] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. 2012. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25 (2012). <https://doi.org/10.48550/arXiv.1206.2944>
- [80] Viriya Taecharungroj. 2023. "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data and Cognitive Computing* 7, 1 (2023), 35. <https://doi.org/10.3390/bdcc7010035>
- [81] Deborah Tannen. 1984. *Conversational Style: Analyzing Talk Among Friends*. Vol. 61. 188 pages. <https://doi.org/10.2307/414501>
- [82] John W. Tukey. 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics* 5, 2 (1949), 99–114. <http://www.jstor.org/stable/3001913>
- [83] Stanford University. 2023. Dialogue distillery: Crafting interpolable, interpretable, and introspectable dialogue from LLMs. In *Alexa Prize SocialBot Grand Challenge 5 Proceedings*. <https://www.amazon.science/alexaprize/proceedings/chirpy-cardinal-dialogue-distillery-crafting-interpolable-interpretable-and-introspective-dialogue-from-lmms>
- [84] Aleksandra Urman and Mykola Makhortykh. 2023. The Silence of the LLMs: Cross-Lingual Analysis of Political Bias and False Information Prevalence in ChatGPT, Google Bard, and Bing Chat. (2023).
- [85] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R Cowan, and Heinrich Hussmann. 2021. Eliciting and analysing users' envisioned dialogues with perfect voice assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15. <https://doi.org/10.1145/3411764.3445536>
- [86] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Büchner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-Based Conversational Agents Using the Psycholinguistic Approach (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376210>
- [87] Sarah Theres Völkel, Ramona Schoedel, Lale Kaya, and Sven Mayer. 2022. User perceptions of extraversion in chatbots after repeated use. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18. <https://doi.org/10.1145/3491102.3502058>
- [88] Jieyu Wang and Anita Komlodi. 2012. Children's Formal and Informal Definition of Technology. In *Proceedings of the 2012 iConference* (Toronto, Ontario, Canada) (iConference '12). Association for Computing Machinery, New York, NY, USA, 587–588. <https://doi.org/10.1145/2132176.2132299>
- [89] Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. *arXiv preprint arXiv:1906.06725* (2019). <https://doi.org/10.48550/arXiv.1906.06725>
- [90] Philip Weber and Thomas Ludwig. 2020. (Non-) Interacting with conversational agents: perceptions and motivations of using chatbots and voice assistants. In *Proceedings of Mensch und Computer 2020*. 321–331. <https://doi.org/10.1145/3404983.3405513>
- [91] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoossa Kasirzadeh, et al. 2021. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359* (2021). <https://doi.org/10.48550/arXiv.2112.04359>

- [92] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023). <https://doi.org/10.48550/arXiv.2302.11382>
- [93] Lu Xu, Leslie Sanders, Kay Li, and James C L Chow. 2021. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. *JMIR Cancer* 7, 4 (29 Nov 2021), e27850. <https://doi.org/10.2196/27850>
- [94] Weilai Xu, Fred Charles, and Charli Hargood. 2023. Generating stylistic and personalized dialogues for virtual agents in narratives. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 737–746. <https://doi.org/10.5555/3545946.3598706>
- [95] Shanshan Yang and Chris Evans. 2019. Opportunities and challenges in using AI chatbots in higher education. In *Proceedings of the 2019 3rd International Conference on Education and E-Learning*. 79–83. <https://doi.org/10.1145/3371647.3371659>
- [96] Jia-Yu Yao, Kun-Peng Ning, Zhen-Hui Liu, Mu-Nan Ning, and Li Yuan. 2023. Llm lies: Hallucinations are not bugs, but features as adversarial examples. *arXiv preprint arXiv:2310.01469* (2023).
- [97] Yee Hui Yeo, Jamil S Samaan, Wee Han Ng, Xiaoyan Ma, Peng-Sheng Ting, Min-Sun Kwak, Arturo Panduro, Blanca Lizaola-Mayo, Hirsh Trivedi, Aarshi Vipani, et al. 2023. GPT-4 outperforms ChatGPT in answering non-English questions related to cirrhosis. *medRxiv* (2023), 2023–05. <https://doi.org/10.1101/2023.05.04.23289482>
- [98] Zhou Yu, Xinrui He, Alan W Black, and Alexander I Rudnicky. 2016. User engagement study with virtual agents under different cultural contexts. In *Intelligent Virtual Agents: 16th International Conference, IVA 2016, Los Angeles, CA, USA, September 20–23, 2016, Proceedings 16*. Springer, 364–368. https://doi.org/10.1007/978-3-319-47665-0_34
- [99] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too? <https://doi.org/10.48550/arXiv.1801.07243> [cs.AI]
- [100] Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. 2022. Large language models are human-level prompt engineers. *arXiv preprint arXiv:2211.01910* (2022). <https://doi.org/10.48550/arXiv.2211.01910>
- [101] Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593* (2019). <https://doi.org/10.48550/arXiv.1909.08593>

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009