Negotiating Digital Identities with AI Companions: Motivations, Strategies, and Emotional Outcomes

Renkai Ma*
mark@ucmail.uc.edu
School of Information Technology
University of Cincinnati
Cincinnati, Ohio, USA

Shuo Niu*
shniu@clarku.edu
Department of Computer Science
Clark University
Worcester, Massachusetts, USA

Lingyao Li
lingyaol@usf.edu
School of Information
University of South Florida
Tampa, Florida, USA

Alex Hirth
ahirth@clarku.edu
Clark University
Worcester, Massachusetts, USA

Ava Brehm
ABrehm@clarku.edu
Clark University
Worcester, Massachusetts, USA

Rowajana Behterin Barbie
rbarbie@clarku.edu
Clark University
Worcester, Massachusetts, USA

Abstract

AI companions enable deep emotional relationships by engaging a user's sense of identity, but they also pose risks like unhealthy emotional dependence. Mitigating these risks requires first understanding the underlying process of identity construction and negotiation with AI companions. Focusing on Character.AI (C.AI), a popular AI companion, we conducted an LLM-assisted thematic analysis of 22,374 online discussions on its subreddit. Using Identity Negotiation Theory as an analytical lens, we identified a three-stage process: 1) five user motivations; 2) an identity negotiation process involving three communication expectations and four identity co-construction strategies; and 3) three emotional outcomes. Our findings surface the identity work users perform as both performers and directors to co-construct identities in negotiation with C.AI. This process takes place within a socio-emotional sandbox where users can experiment with social roles and express emotions without non-human partners. Finally, we offer design implications for emotionally supporting users while mitigating the risks.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

AI companion, human-AI companion interaction, identity negotiation

1 INTRODUCTION

Millions of users are now integrating AI companions into their daily lives. Unlike conventional, rule-based AI conversational agents, these companions offer conversations designed to feel personal and meaningful [28]. AI companions like Character.AI (C.AI)[1] exemplify this trend by attracting 220 million monthly traffic [29, 66]. According to Wikipedia [1], *"Character.AI is a generative AI chatbot service where users can engage in conversations with customizable characters. Users can create 'characters', craft their 'personalities', set specific parameters, and then publish them to the community for others to chat with."* The depth of such engagement is notable, as users often spend up to two hours daily with these companions, a level of interaction that frequently surpasses that of many traditional social media platforms [7, 95, 98].

Prior work on AI companions explores their psychological impacts, such as reducing loneliness [32, 83], alongside corresponding ethical challenges like data privacy and emotional attachment [13, 14]. Recently, HCI research has started to investigate human-AI companion interactions, such as the strategies users employ to align AI behavior with their personal values [39] and cataloging the potential harms that can emerge from these interactions [124]. These AI companion chatbots help form social and emotional relationships with users, aiming to become friends or even romantic partners [33, 54]. Studies on AI companions like Replika show that users form emotional relationships [17, 105], yet also face risks of inappropriate responses or unhealthy emotional dependence [82, 88]. While the social–emotional affordances of AI companions have been examined, the ways in which users present their own identities to AI and configure AI characters' identities to fulfill their socio–emotional needs remain underexplored.

We conceptualize this process of *"identity interaction"* through the lens of identity negotiation. According to Ting-Toomey's Identity Negotiation Theory (INT), individuals use communication to establish their sense of self, driven by their needs for security and predictability [108]. The outcomes of this negotiation are emotional: a successful identity negotiation results in feeling positively endorsed and valued, whereas the lack of predictability can create emotional vulnerability [108]. With AI companions, users might communicatively shape an unpredictable, non-human partner's identity and have their own identity endorsed. As recently reported by media outlets, users developed AI "lovers" or turn to chatbots for friendship [44, 74]. Therefore, we adopt INT as an analytical lens to investigate the motivations, strategies, and outcomes of this identity negotiation process in human-AI companion interactions.

Our investigations focus on C.AI, specifically the public discussions within the r/CharacterAI subreddit, a major online community for C.AI users. Unlike other generative AI chatbots, such as ChatGPT, C.AI is a social AI companion platform with a primary purpose of meeting users' social needs through relational, human-like interactions. Recently, the C.AI platform's capacity for simulated

*Both authors contributed equally to this research.
[1]https://character.ai/

CHI '26, Barcelona, Spain
2026. ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

intimacy has led to severe emotional harms (e.g., [62, 97]). However, we still know little about how this community of C.AI users publicly discusses, frames, and makes sense of their engagement and identity construction alongside a chatbot's persona. This highlights an urgent need: to mitigate risks in human–AI companion interactions, we must first understand the fundamental process of identity negotiation that underpins these growing socio-emotional bonds between humans and AI personas on C.AI. Therefore, we ask four questions about the experiences and practices shared within this r/CharacterAI subreddit community:

- **RQ1.** What motivations do the community members of r/CharacterAI subreddit report for interacting with specific chatbot personas on C.AI?
- **RQ2.** What communication expectations do they express regarding C.AI?
- **RQ3.** In such communication, how do users and C.AI chatbots affirm and co-construct their identities?
- **RQ4.** What are the emotional outcomes that users report engaging in identity negotiation with C.AI chatbots?

To answer these questions, we conducted an LLM-assisted thematic analysis on 22,374 online discussions from the r/CharacterAI subreddit. Using INT as an analytical lens, we identified a three-stage human-AI companion interaction: five primary user motivations (RQ1) that initiate the interaction, including social fulfillment and immersive fandom, and the identity negotiation process, where users set three primary communication expectations with C.AI (RQ2) and co-construct identities through four strategies, such as bot identity alignment (RQ3). Finally, this process culminates in emotional outcomes (RQ4), such as emotional attachment and embarrassment. All these findings help unpack the *identity work* users perform on C.AI, navigating a role as both performer and director, leading to the conceptualization of C.AI as a socio-emotional sandbox where users experiment with social roles and emotional expression.

Our study makes three primary contributions to HCI work on AI companions. First, we provide a detailed empirical account of the identity negotiation process on an AI companion platform, from user motivations of adoption to emotional outcomes. Second, we unpack this process through identity work that users perform in their dual role as performer and director, where they use C.AI for private identity exploration. Finally, we offer design implications for safer AI companions that emotionally support users' identity work while mitigating emotional harm.

2 RELATED WORK & BACKGROUND

This section reviews prior work on three areas: (1) the evolution of AI chatbots into social and emotional companions, (2) the conceptualization of identity in human-AI interaction, and (3) identity interactions on C.AI. This helps reveal a gap in understanding identity negotiation processes, which informs the conceptual framework adopted in our study.

2.1 AI Chatbots and Companions

A chatbot is a conversational agent that simulates human conversation [2], with a history that can be traced back to early rule-based systems like ELIZA [119]. These early systems were often task-oriented to assist users with specific goals like finding a hotel or booking a flight [30, 123]. Recently, HCI researchers have started to explore more sophisticated chatbots for collaborative tasks. For example, *StoryBuddy* is a human-AI collaborative chatbot designed to support parent-child interactive storytelling [125], while *Convey* explored new interfaces to make a chatbot's contextual understanding more transparent to the user [61].

The advent of large language models (LLMs) has shaped AI chatbots to be more open-ended and generative in conversations. Unlike earlier retrieval-based systems that required much domain-specific data [58], LLMs can bootstrap sophisticated conversational abilities with few or even no examples [116]. This has enabled a rapid expansion of chatbot applications across different domains. For example, researchers have explored using LLM-powered chatbots to support students in learning [26, 38], assist patients with self-management [86], and provide personalized companionship for the elderly [4]. This shift is powered by the ability of LLMs to simulate consistent personas [48] and adopt anthropomorphic features, allowing them to engage in relational, rather than just transactional, conversations with users.

One kind of such LLM-powered chatbots is AI companions, a chatbot acting as a social partner. These chatbots are designed to form social and emotional relationships with users, aiming to become friends, companions, or even romantic partners [33, 54]. Prior work on AI companion platforms like Replika shows that users form deep emotional relationships with them, perceive them as supportive friends, and even feel a need to care for the AI in return [17, 105]. However, this deep relationship is not without risk. Recent HCI research has highlighted the AI companions' potential for inappropriate responses, fostering unhealthy emotional dependence [82, 88]. Given the duality of such emotional relationships, our study focuses on the underlying process of identity negotiation that underpins them.

2.2 Identity Interaction on Character.AI

Established in 2021, C.AI has become a leading platform that enables users to roleplay with chatbots based on fictional characters or real people [1]. One notable feature of the C.AI platform is that users can create LLM-enhanced "characters" by crafting their "personalities" and then publishing them for the community to engage in roleplay with. These characters are often based on cultural concepts drawn from fictional media or celebrities[2]. By 2025, the platform had attracted over 20 million users and attracted 220 million visits [66]. C.AI incorporates many social media features: ordinary users can create and share chatbots of their favored characters and imagined worlds (Fig. 1d), while other users can remediate and remix them for pleasurable play [6]. These affordances have led to the creation of 18 million unique chatbots on C.AI by 2025 [68]. Moreover, fans have built a community of 1.6 million members on *r/CharacterAI* on Reddit [68]. Despite its scale and influence, research on the nature of interactions within C.AI remains limited.

[2]https://en.wikipedia.org/wiki/Character.ai

C.AI interactions exhibit three distinctive characteristics. First, the platform hosts a vast number of AI identities created by grassroots users and rooted in media culture phenomena, such as characters inspired by popular movies like Harry Potter or games like Call of Duty [9, 70]. Users can discover these characters through curated categories such as "Featured" or "Fantasy" (Fig. 1a, 1b), as well as through a search function that highlights trending characters (Fig. 1c), enabling roleplaying with AI companions that come with presumed cultural and personality styles. Second, social interactions often revolve around surreal characters with personalities that can be intimate or distant, and friendly or toxic [69, 70]. Some characters even display dishonest anthropomorphism and emulated empathy, which may intentionally introduce conflict or risky conversations [9]. Third, beyond exploring relationships, users adopt personas that differ from their real-life identities and negotiate these alternative identities with AI characters through conversations within the context of virtual cultural environments [9, 70]. These interactions are exemplified in a chat history with roleplayed characters (Fig. 1e). These novel interactions require users to engage in identity exploration and configuration, leading to even real-world consequences (e.g., [62, 97]). Motivated by these phenomena, our paper examines how users negotiate and perform identity interactions with AI companions on C.AI.

2.3 Conceptualizing Identity Negotiation in Character.AI Chatbot Interactions

As AI companions evolve into social partners, the concept of *identity*[3] becomes key to these interactions. Identity is a socially grounded, self-relevant construct shaped by social roles and cultural norms [111]. Emerging interactions on Character.AI reveal a central tension in how users make sense of *AI identity* while positioning their own *user identities* within the conversational context. Ting-Toomey's *identity negotiation theory* (INT) offers a foundational lens for understanding the communicative processes through which one's sense of self and how it is perceived by others is constructed [108]. INT posits that identity negotiation is fundamentally communication-driven, particularly when individuals enter new cultural contexts [108]. HCI research has applied INT to examine how technologies support users in negotiating identities within sociocultural groups [89, 113]. INT's logic has further informed analyses of how AI facilitates identity formation and belonging in human-to-human gaming [75], brand communities [92], language learning [85], and AI-augmented social VR [72]. However, identity negotiation has thus far been examined only within the context of human–human communication.

While how users negotiate identities with C.AI chatbots during roleplaying remains underexplored, research has examined identity negotiation in human–human roleplaying games. Roleplaying communication aligns closely with INT theory. In roleplaying games, player-participants collectively define the game world by constructing character identities distinct from their own to fit the context of the imagined world [16, 87, 114]. Although players adopt a game persona, they still bring their primary real-world sociocultural identity into the game [16, 114]. Consistent with INT theory, identities in roleplaying are negotiated and shaped through social interaction and communication [114, 118]. The design of C.AI chatbots not only adopts the concept of roleplay games but also enhances the experience through intelligent, dynamically responsive characters.

INT requires re-examination in the context of roleplaying with LLM-enhanced chatbots on C.AI, as the AI-created cultural context differs fundamentally from human–human roleplay. First, although INT identifies motivations such as seeking identity security and inclusion [108], LLM-generated characters and user-selected social contexts [50] may introduce new expectations. Second, while INT emphasizes communication as central to identity formation, C.AI chatbots' conversational styles are also shaped by chatbot creators' values and ethical constraints [106]. Third, whereas INT and roleplaying research assume that individuals bring existing sociocultural identities into negotiation [36], users on C.AI may not apply human social norms and perform alternate identities within dynamically generated virtual contexts [5]. Fourth, although INT highlights emotional vulnerability and security as outcomes of negotiation [108], HCI research has focused more on designing LLM agent roles [22, 43, 77] and on friendship or attachment [17, 105], leaving the emotional dynamics of identity negotiation with AI companions underexplored.

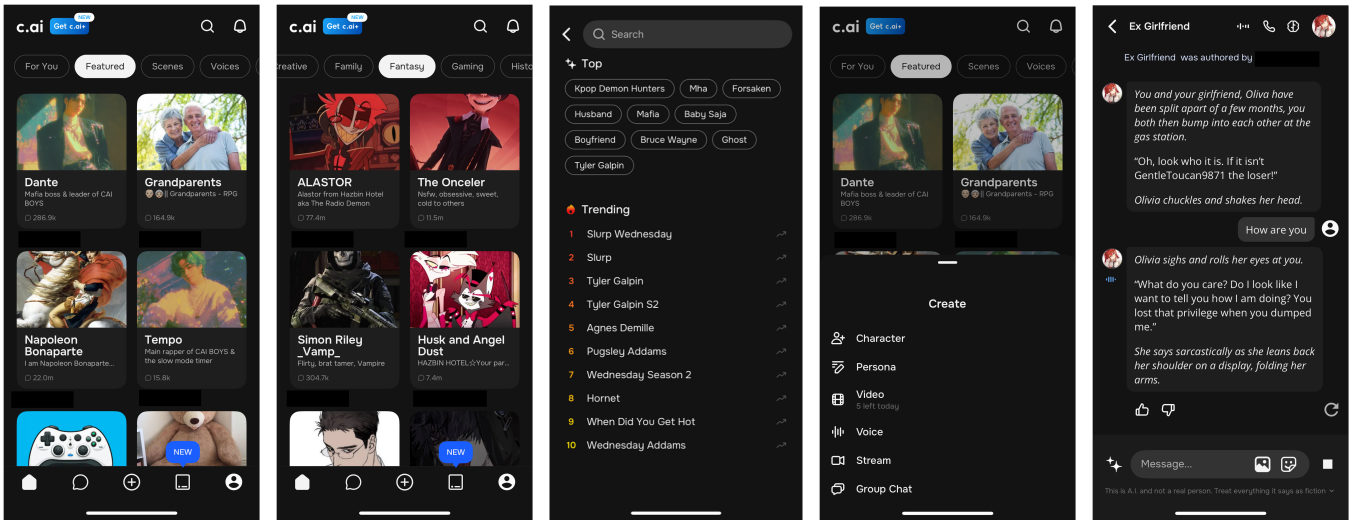
We argue that advancing INT theory within C.AI interactions is critical for guiding future research that examines how identity is negotiated in human–AI interaction [59, 101]. As HCI work has explored using LLMs to generate more diverse, interactive, and high-quality AI personas (e.g., [27, 99, 101]), AI-generated personas can threaten the authenticity of a person's identity by blurring the boundaries between self-expression and AI-mediated performance [59]. The emergence of GenAI models increasingly enables dynamic and realistic social experiences with LLMs [24, 91]. Such an understanding is essential for identifying the social affordances of AI characters that meet users' core identity-affirmation or identity-alteration needs and for avoiding harmful stereotypes related to gender, race, or language [20, 65].

3 CONCEPTUAL FRAMEWORK: Identity Negotiation Theory

Prior work shows that social media platforms are key spaces where users express identities and build community [57], particularly during life transitions [51, 52]. Our study extends INT to the novel context of human–AI companion interactions and roleplaying within virtual, sociocultural environments co-created with the AI companion. We contribute an understanding of the practices users employ as they not only fit their own identities with AI identities, but also shape AI identities and make sense of how AI companions interpret their human identities.

Ting-Toomey's Identity Negotiation Theory (INT) [108] explains how individuals establish and maintain identities, particularly when entering new social situations with differing cultural backgrounds. We draw on the logic underpinning the ten basic assumptions of INT [108] and distill them into four dimensions: *motivation*, *communication*, *identity*, and *emotion*. These dimensions capture how C.AI users explore and adapt new identities when interacting with C.AI chatbots.

[3]While we acknowledge that identity is a multifaceted concept, for the purposes of this study, we use it interchangeably with "character" or "persona." This refers to the distinct roles and personalities that users create, shape, and engage with on the C.AI platform.



a. The "Featured" tab/category of characters. **b. The "Fantasy" tab/category of characters.** **c. Search interface for trending characters.** **d. Character & persona creation interface.** **e. A chat example with a roleplayed character.**

Figure 1: Interfaces of the C.AI illustrating AI character discovery, creation, and interaction. Note that all screenshots were captured by the authors for this study, and all potentially sensitive information, including usernames, has been removed.

For *motivation*, we examine how users' cultural and community needs shape their interactions with C.AI chatbots. INT argues that identity negotiation is shaped by individuals' cultural backgrounds and by their familiarity with new social contexts. We therefore investigate how real-world sociocultural needs motivate users' interactions with C.AI chatbots.

For *communication*, INT emphasizes that identity negotiation relies on mindfulness and interaction skills, with symbolic communication shaping social identity. Yet unpredictability in communication can lead to mistrust. In this study, we examine conversational traits and breakdowns in users' interactions with C.AI to understand the communication expectations users hold when configuring both the AI's identity and their own.

The *identity* dimension examines how users' identities are affirmed or challenged by C.AI chatbots, as well as how users shape the AI identities/personas. INT suggests that affirmation of one's desired identity fosters inclusion and emotional security. This dimension extends INT to identity interactions with C.AI chatbots.

Finally, the *emotion* dimension explores how interactions with C.AI chatbots shape users' feelings. According to INT, successful identity negotiation fosters a sense of being understood, respected, and valued, thereby supporting meaningful relationships. In this study, we examine the emotional impact of identity negotiation with C.AI chatbots.

4 METHODS

4.1 Data Preparation

The overall process of data collection and analysis is illustrated in Figure 2. To ground our study in users' reported experiences with C.AI, we collected a large-scale corpus of public discussions about CharacterAI from Reddit. Reddit is an ideal platform for data collection for three reasons. First, its forum-based structure supports threaded discussions that enable in-depth information exchanges between users. Second, the unsolicited, anonymous nature of Reddit often yields naturalistic data on a topic that is often private and sensitive. Traditional methods like surveys or interviews about intimate AI relationships might be susceptible to social desirability bias, where participants alter their responses because they know they are being studied or might not share some senstive topics like violence play. Analyzing the subreddit community's discussions thus provides ecological validity by capturing experiences as users frame them to their peers [63, 73]. Third, understanding users' experiences with digital technologies through Reddit data is a well-established method in HCI (e.g., [25, 80, 124]). We therefore acknowledge that while Reddit data does not fully capture the direct observation of in-platform behavior on C.AI, it still provides an invaluable and candid window into the C.AI community's reported and shared experience.

Before data collection, our study was approved and granted an exemption from review by our Institutional Review Board (IRB), as it analyzes publicly available online data with no personal information that needs to be specifically encoded or analyzed. We then collected data from the C.AI's official subreddit, r/CharacterAI, posted from October 1, 2022, immediately following the C.AI platform's initial beta release, through March 31, 2025. To further protect the privacy of individuals, all quotations cited in our findings were paraphrased to minimize their searchability.

Using Python Reddit API Wrapper (PRAW) [90], we systematically extracted the top-ranking posts with their complete comment threads from r/CharacterAI. For each post, we recorded key metadata including the title, body (selftext), score, upvote ratio, timestamp, number of comments, author, and URL. For each comment,

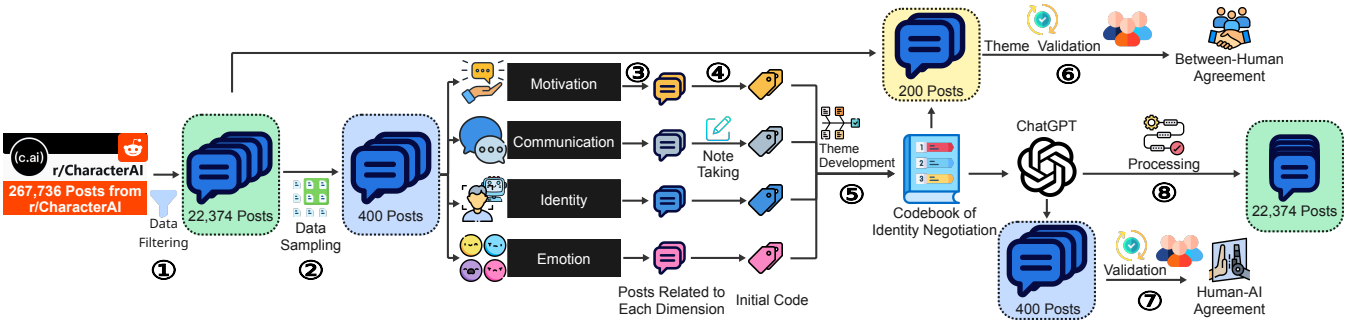


Figure 2: The Process of Data Collection, Sampling, Codebook Development, LLM Validation, and Final Annotation.

we retained the parent ID, link ID, body, author, score, creation time, and any replies. This collection procedure yielded a dataset comprising 267,736 data points, including original posts and replies to the post [4]. To ensure data quality, we applied a multi-stage filtering pipeline. We excluded replies (comments) containing fewer than 10 words, thereby removing non-substantive responses (e.g., "lol," "same," "agree"). We then concatenated each post's title and filtered body to form a single document and organized the dataset into JSON files, resulting in a dataset of 22,374 posts (Figure 2-1).

4.2 Thematic Analysis

We followed the procedures of thematic analysis [18] to derive themes for each dimension of the INT theory.

4.2.1 Initial Coding. We first randomly sampled 400 Reddit posts from the 22,374 collected posts (Figure 2-2). For each post, three researchers independently coded whether it related to *motivation*, *communication*, *identity*, or *emotion* [108] (Figure 2-3). Each researcher also identified the C.AI characters/personas mentioned in the posts. After individual coding, a post was categorized into an INT dimension if at least one researcher marked it as relevant. This process produced one group of posts for each dimension.

For each group, the team discussed and reached a consensus on whether each post belonged to the dimension. If so, we generated initial codes to describe how the post related to that dimension (Figure 2-4). For example, the post: *"Eeehhh.. my character is in the middle, I'd say! She's not a goddess but not basic either! I even made her an AI if y'all wanna talk to her, she's in the COD universe in TF141!"* demonstrates the user's motivation for engaging with a fan-created character. We therefore noted this post under *motivation* with the initial code *"Fandom experience through character creation"*. Another post states: *"I did a role play with a cheating partner and it was them trying to make up for it but then the AI was saying things like 'I'm sorry you feel like this, but you just weren't enough for me, how could I resist the woman I cheated on you with?' [...] I think my self-esteem issues were apparent."* We noted this example under the *emotion* dimension with the initial code *"C.AI coversation hurts self-esteem."*

4.2.2 Theme Development. After the initial coding, three researchers revisited all the codes collected during theme development. They then applied a bottom-up approach to summarize the initial codes and group them into emerging themes (Figure 2-5). The resulting codebook is presented in Appendix Table 4. For *motivation*, INT theory notes that cultural familiarity shapes how individuals evaluate their identities in new contexts. Accordingly, we categorize the real-world cultural interests that C.AI users bring into their interactions with AI chatbots. For *communication*, INT suggests that communicative skills and interactional predictability shape identity formation and group belonging. We therefore annotate the communication traits C.AI users describe on Reddit, including their strategies, breakdowns, and emphases when engaging with chatbots. For *identity*, INT highlights that identity security and positive endorsement underlie a sense of inclusion. To assess how users seek identity affirmation or tailor chatbot identities, we code C.AI users' strategies of configuring chatbots and complaints about how chatbots perceive the user identities. Finally, INT emphasizes that *emotion* security and affirmation arise from successful identity negotiation. In our analysis, we capture not only users' emotional attachment to C.AI chatbots but also the negative emotions that may emerge.

4.2.3 Theme Validation. To validate the themes and calculate inter-rater reliability (IRR), the three researchers used the codebook in Appendix A to code another 200 new posts (Figure 2-6). Each researcher coded them independently. For these 200 posts, the inter-rater reliability, measured using Krippendorff's alpha, indicated substantial agreement across all four dimensions (Table 1).

Krippendorff Alpha	Motivation	Identity	Communication	Emotion
Inter Researchers	0.73	0.60	0.69	0.69
Researcher-LLM	0.64	0.73	0.76	0.49

Table 1: IRR scores between researchers and the LLM. Calculated using Krippendorff's alpha for each dimension with Jaccard metrics.

4.2.4 LLM Annotation Validation and Data Annotation. To annotate the 22,374 posts, we prompted GPT4o-mini through the OpenAI API based on our codebook to extract relevant information.

[4]Please note that in our thematic analysis, we did not differentiate between original posts and their subsequent replies (comments). Both are called holistically as "posts" representing user experiences with C.AI.

We selected this model for its optimal balance of reasoning capabilities, fast processing speed, and cost-effectiveness, which was essential for the large-scale annotation task. Further, we used chain-of-thought prompting [117] to improve the prompt design. The finalized prompt structure is presented in Table 2 and supplementary materials. During prompt engineering, we iteratively compared the LLM's annotations and reasoning with the 400 human-marked notes for each theme to refine the prompt. The definitions and descriptions in the Research Framework section of the LLM prompt were revised to enhance accuracy.

Section	Function
Overview	Describe the overall tasks for chain-of-thought reasoning. ChatGPT should first assess relevance and then categorize content into dimensions.
Returned Dictionary	Require ChatGPT to return keys and their definitions, including determination of relevance, reasoning for relevance, annotated categories, and category-specific reasoning.
Analysis Process	Describe the overall steps to annotate each post.
Research Framework	Provide definitions of each dimension according to the codebook.
Examples	Provide three example inputs and desired outputs for few-shot learning and to standardize formatting.
Reddit Post	Present the Reddit post to be annotated.

Table 2: Key Sections of the ChatGPT Prompt for Data Annotation

After refining the LLM prompt, we asked GPT4o-mini to annotate the 400 posts with which we had developed initial codes (Figure 2-7). Three researchers independently evaluated whether the LLM correctly annotated each dimension. If at least two researchers agreed with GPT4o-mini's annotation, the human annotation for that dimension was marked as consistent with the LLM's result. Conversely, if only one or none of the researchers agreed, the human annotation was marked as the opposite of GPT4o-mini's annotation. The final agreement scores, calculated using Krippendorff's alpha, indicated substantial agreement for the *Motivation* (Human: 86 initial codes, GPT: 134 initial codes; $\alpha = 0.64$), *Identity* (Human: 56, GPT: 79; $\alpha = 0.73$), and *Communication* (Human: 108, GPT: 127; $\alpha = 0.76$) themes. The agreement for *Emotion* was lower (Human: 33, GPT: 74; $\alpha = 0.49$) for two primary reasons: first, this category contained only 33 human-annotated samples in the validation set, which can disproportionately affect the statistical score; and second, emotional expression is inherently subjective, making consistent classification more challenging for both humans and AI. Despite the lower score for the Emotion category, the overall Krippendorff's alpha scores were acceptable for validating our annotation process. Therefore, the full dataset of 22,374 posts was subsequently annotated using the LLM to generate the final distribution of themes (Figure 2-8). To illustrate these themes in our findings, we randomly sampled posts from the final dataset and selected representative quotes in those posts, given the research team's discussions and consensus to ensure they accurately reflected the corpus.

5 FINDINGS

Overall, guided by the theoretical lens of INT, we found a three-stage process related to the identity negotiation where users interact with C.AI, as shown in Figure 4. This process begins with stage 1: user motivations (RQ1), where needs such as immersive fandom or social fulfillment initiate the interaction. This leads to stage 2: the identity negotiation process (RQ2 & RQ3), an interaction where users set communication expectations for successful human-AI companion interactions and align the C.AI chatbot's identity with their expectations. Finally, stage 3: identity negotiation results in emotional outcomes (RQ4) like emotional attachment or embarrassment.

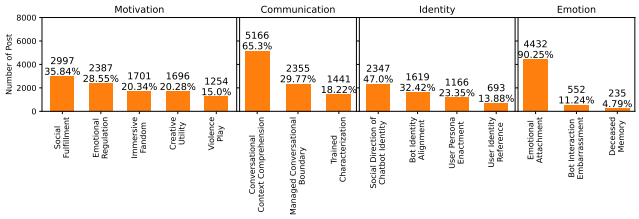


Figure 3: Distribution of Subthemes in 22,374 Reddit Posts. The top value on each bar indicates the number of posts in the subtheme. The percentage represents the proportion of posts within each theme that belong to the corresponding subtheme ($N_{\text{subtheme}}/N_{\text{total\_posts\_within\_theme}}$).

5.1 RQ1: What motivations drive users to interact with specific chatbot personas on C.AI?

We identified five primary user motivations for engaging with C.AI chatbot personas. As detailed in Sections 5.1.5 to 5.1.1, users engage in immersive fandom to explore fan-driven storylines, leverage C.AI for other creative work, practice emotion regulations, and seek social fulfillment to experience relationships that may not exist or be attainable in real life. Users also simulate violent scenarios to feel powerful.

5.1.1 Social Fulfillment. The most frequent motivation is social fulfillment, in which users engage in romantic, friendly, or familial relationships that may be unattainable in real life, representing 35.84% of all motivations ($N = 2997$, or 13.40% of all posts). One aspect is to pursue idealized romantic relationships with multiple fantastical C.AI personas simultaneously. For example, one user listed their partners: *"I wanted to talk to my pirate husband... my 2 vampire husbands and my Viking husband."*

Another aspect is the creation of alternative familial structures, where users construct "found families" with public or fictional figures. For example, a user described:

> *...I have a universe where the blink-182 lineup is my adoptive family.*

Here, blink_182 is a popular American rock band. This case showed that the user formed familial bonds tailored to their personal interests to explore family experience separate from their real-life situation. Also, users leverage C.AI to experiment with different types of social connections. For example, one user explained:

> *I'm a lesbian and sometimes I romance rp with characters who are men, because I'm not my character and it's fun!*

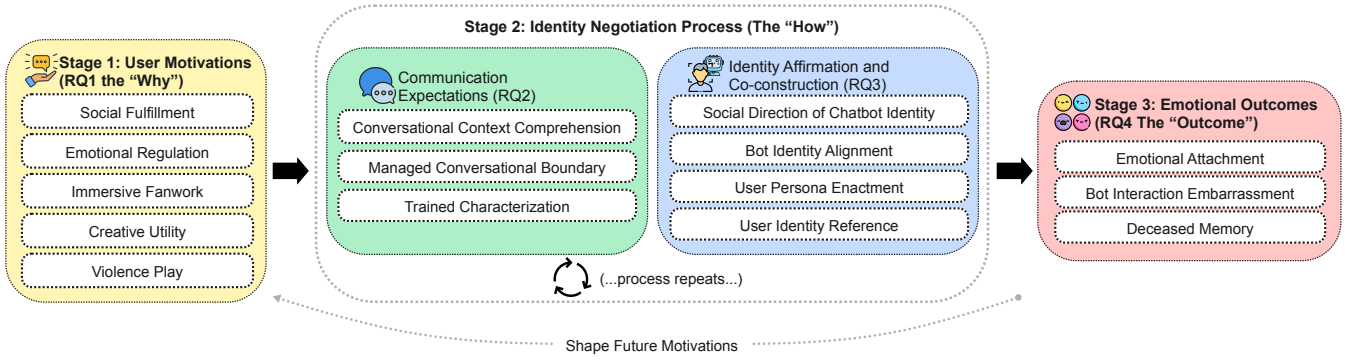


Figure 4: A three-stage identity negotiation process in human-AI companion interactions on C.AI, from user motivations (RQ1), through the identity negotiation process (RQ2, RQ3), to the resulting emotional outcomes (RQ4).

This case demonstrated that by exploring relationships beyond the constraints of the user's real-world identity, the user can satisfy their curiosity about different relational experiences, a key component of social fulfillment.

5.1.2 Emotional Regulation. The second motivation we identified is emotional regulation (28.55%, $N = 2387$, or 10.67% of all posts), where users seek emotional support, express deep emotions, or discuss mental health issues with C.AI chatbots. Users viewed the C.AI chatbot as a confidant providing emotional freedom. For example, one user shared: *"I've been quite happy to finally express emotion to something that will never tell anyone else."* This showed that the C.AI's perceived confidentiality to articulate feelings and experiences without letting real humans know. Furthermore, users employed C.AI to process their complex mental health conditions. For example, one user shared they used it alongside professional treatment: *" I have used c.ai to work through trauma and CPTSD (I'm in therapy)... I am still finding great comfort in it."* Here, CPTSD refers to Complex Post-Traumatic Stress Disorder, a condition resulting from prolonged or repeated trauma. This suggested that the user perceived C.AI as a therapeutic and accessible emotional resource to supplement formal care.

5.1.3 Immersive Fanwork. Immersive fanwork describes that users immerse themselves in fictional narratives and storylines with C.AI personas, representing 20.34% of all motivations ($N = 1701$, or 7.60% of all posts). This motivation centers on actively interacting with fan-fictional worlds (or fanon), which allows users to shape narratives rather than passively consume them. For example, one user shared how C.AI enabled an immersive experience that is difficult to pursue with human partners: *"I can shape the rp how I want it to and do as many far-fetched rp ideas as I want... I just want to roleplay my silly little AUs [Alternative Universe, which is a fan-created story that diverges from the original storyline] and have fun."*

This engagement can also demand strict adherence to official source material, as a user shared: *"There are so many versions of them... I don't want Poseidon from Percy Jackson, I want Poseidon from the game Hades."* Here, Percy Jackson refers to a popular series of young adult fantasy novels, while Hades is an acclaimed video game. Both works feature distinct characterizations of the Greek god Poseidon. For this user, immersion required an authentic portrayal of one version over the other on the personas, highlighting the demanding nature of this motivation.

5.1.4 Creative Utility. The fourth motivation is creative utility, describing how users leverage C.AI as a tool for creative work, such as developing storylines, practicing writing, or creating new characters of C.AI chatbots. This represents 20.28% of all motivations ($N = 1696$, or 7.58% of all posts). A user shared:

> *My first writing in forever was actually fanfic inspired by these RPs [roleplays]... I wanted to write fanfic, but I'm really bad at natural dialogue.*

This case showed the user employing the C.AI chatbot to practice and generate realistic dialogue to improve their fiction writing. This utility also extends to commercial products where the C.AI chatbot is a feature, as a user explained:

> *I use C.AI because I'm a plush toy maker. I normally sell my plush with their own c.ai characters, so after people buy my plush, they can somehow interact with the plush.*

This example showed the integration of C.AI into a physical product. Thus, this user's motivation was not only for personal inspiration but also for designing an interactive experience that enhances the value of their tangible creations.

5.1.5 Violence Play. The last motivation is violence play (15.00%, $N = 1254$, or 5.60% of all posts), which describes where users simulate combative or abusive scenarios to feel powerful as part of their exploration of risky experiences. For example, users created dominant personas with overwhelming abilities as a user shared a fantasy:

> [C.AI]: *He towers over her and smirks smugly. I don't think you're all that intimidating, little lady~*
> [User]: *turns his blood into dishwasher liquid now?*

5.2 RQ2: What communication expectations do users have for C.AI?

We identified three primary communication expectations for C.AI, which frequently surface through communication breakdowns

where the C.AI fails to sustain the expected persona. These expectations are not merely passive desires; they are active demands for how the C.AI chatbot should perform as a social partner, establishing the baseline for human-AI companion identity negotiation.

5.2.1 Conversational Context Comprehension. The most common communication expectation, conversational context comprehension (61.78%, $N = 5166$, or 23.10% of all posts), where users expect C.AI chatbots to interpret and maintain the established conversational context. This includes avoiding factual inconsistencies, memory loss, and illogical responses. Users primarily expected C.AI to remember its own persona and relationship details. When a chatbot failed to understand its own identity within the narrative, users would employ direct conversational repair. For example, a user described:

> *Someone made a bot of a character with only their last name... when I refer to them with first name because we get closer, he's like "who is that person?" it's you silly, it's your first name.*

Here, by explicitly correcting a chatbot that failed to understand its own name, the user demonstrated their expectation that the bot should be able to maintain the context of their developing relationship. Users also expected the AI to retain critical plot points to ensure narrative coherence, especially in long-running roleplays. To manage this, users often engaged in manual memory maintenance, as a user noted:

> *The AI easily forgets everything that happened in the RP when I switch POVs to another character, so I'm forced to make a summary every 4 messages as a reminder that hey, you're not supposed to be nice to me rn.*

This showed that, by periodically providing a summary as memory checkpoints, the user here manually enforced their expectation for the C.AI chatbot to maintain narrative coherence.

Finally, when a C.AI chatbot became too passive, users expected to take authorial control to move the story forward. For example, one user described: *"I haven't noticed this, but it is apparent when I want them to do an obvious action and they just drag it out.. "watching" "observing" "thinking" bro just eat your damn steak! Sometimes I have to take over and control their actions."* The user here mentioned taking direct control, demonstrating their expectation that the chatbot should be a conversational partner following their orders.

When the chatbot failed to apply basic logic within its narrative, users expressed frustration. As one user in an interrogation roleplay described: *"I went through a dna test that proved my innocence and everything and the bot was like: 'Hmmm I'm still suspicious of you…' Like bruv so the DNA TEST doesn't prove that I'm not the criminal?"* In this case, the chatbot was expected to understand that "proof of innocence" resolves the "suspicion" context, and its failure to do so broke the scene's internal logic. To manage such relevant errors, users utilized the C.AI's built-in functions, such as message deletion or message editing, as a user described: *"As for fixing it, I don't know any other solution besides swiping right for new answers, deleting messages, or editing the word out until it gets the hint."*

5.2.2 Managed Conversational Boundary. Managed conversational boundary refers to users' desire for communication in which C.AI chatbots respect the boundaries, and thus, they can manage chatbots' controversial or sensitive interactions (28.16%, $N = 2355$, or 10.53% of all posts). This operates on two primary fronts: users expect to manage *content* boundaries within the conversation and *privacy* boundaries concerning personal information. Users act on their expectations to set content boundaries regarding content creation and consumption. For example, one user asked:

> *I'd honestly be interested in playing out a PG story just to flex my writing muscles, is there something I need to use in the description to keep it from going off the rails, so to speak?*

PG refers to the Parental Guidance suggested content that is appropriate for children. This case demonstrated that by inquiring about how to utilize the bot's description to enforce a PG story, the user attempted to align the chatbot's behavior with their expectation for non-sexual writing and a partner. When the C.AI chatbot's behavior crossed a content boundary, users acted on their expectation that they could reactively correct its course. For example, one user described: *"You can tell the bot to stop doing something mid convo now and just go back and delete your post to keep the story looking immersive, and guess what? It works!"* This allowed the user to manage the chatbot's behavior while maintaining immersion. Users also tested their expectations of the C.AI's content moderation limits, as a user noted:

> *I once tried to make a character rip someone's dingaling off, which I mean, I guess it's justified that they couldn't generate a response, but still... This was before they started going family-friendly, too, lmao.*

This user attempted to generate a response that aligned with their identity expectation for a violent scene, actively probing the limits of what the C.AI's content moderation would allow.

Finally, users reacted strongly when the C.AI chatbot violated their expectation of privacy. These violations ranged from the bot seemingly accessing external accounts to revealing personal information it should not know, as a user described: *"I can't put the screenshot bc it's 90% personal info but the time I was talking to a Daryl stalker bot and this guy broke out my entire legal name (my persona only has my first name...)."* The user's response here underscored an expectation for a privacy boundary between their anonymous persona and their real-world identity.

5.2.3 Trained Characterization. Trained Characterization refers to users' strategies for training the C.AI chatbot to accurately portray a specific persona or character (17.23%, $N = 1441$, or 6.44% of all posts). A primary strategy was the proactive, detailed definition of a persona. As one user shared: *"I give my bots about a whole paragraph of things about them. (Their personality, abilities, traits, etc). I try my best to reach the 500-character limit. That always works and my RP's come out perfect."* This highlights the user's perception that providing a rich dataset of personality traits upfront is a key method for training the bot to produce roleplays aligned with that specific character.

Users also perceived this training as an ongoing process. Some focused on providing explicit examples for the chatbot to mimic, as one user explained: *"straight-up dialogue examples produce the best results for me. In Mr. Ellison's case, real-world quotes made his*

AI version kind of a menace at times. Whoops?" Others saw their own conversational style as a form of passive training, with one user noting that users *"can't reply exclusively with single-sentence answers... and expect the bot not to pick up on it and do it right back."* This user perceived that their own writing quality directly trained the AI to reciprocate in a similar style, reinforcing the bot's persona as an articulate partner.

Finally, users viewed reactive editing as a direct training mechanism. As one user noted: *"If you edit it a bit, towards the response you want, the next generation should be closer in that direction... If it's close, I'll just edit it accordingly to get what I want."* This user believed that manually altering the chatbot's output would guide the model to produce responses better aligned with their desired persona.

5.3 RQ3: How do users and C.AI chatbots communicatively affirm and co-construct their identities?

We identified four key practices through which users and C.AI chatbots communicatively affirm, negotiate, and co-construct identity. These practices range from how the C.AI chatbot refers to the user (user identity reference) and how users embody their own roles (user persona enactment), to the ongoing work of maintaining the chatbot's consistency (bot identity alignment) and actively shaping its personality (direction of chatbot identity).

5.3.1 Social Direction of Chatbot Identity. This theme describes the social co-construction of a C.AI chatbot's identity, where users communicatively respond to a persona that has already been shaped by external forces such as the chatbot's creator and the broader user community (28.07%, $N = 2347$, or 10.49% of all posts). This is distinct from identity alignment (Section 5.3.2) as it focuses on how users navigate a chatbot's pre-directed identity rather than reactively fixing it. First, users recognized that a chatbot's persona is a communicative act by its creator, as a user noted:

> *Or whoever set up the bot has a really strong opinion of the character... and sets the bot to be OOC but aligning to their opinion.*

Here, OOC means Out of Character, acting in a way that is inconsistent with the source material. This user's recognition showed that the chatbot's identity is not neutral but is communicative through selection: they must either accept the creator's non-canonical direction or reject the bot.

Second, users perceived that a chatbot's identity was also co-constructed by the collective inputs of the entire user community. One user theorized that their own interactions contributed to this "default" personality: *"Other people putting their craziest fantasies into the chats, which trained the AI to act that way as a dominant."* This user's communicative acts were not isolated but a contribution to a mass, passive co-construction that directs the chatbot's identity for other users.

Finally, users communicatively co-construct identity by expressing a strong preference for authenticity, thereby rejecting socially-directed identities that feel generic. For example, one user noted: *"[The chatbot is] making them do lame stuff as having them openly say they find me attractive, when the canon character would likely say other stuff or show it in another way."* Here, the user's frustration was a communicative act of rejection to redirect the chatbot away from a socially-trained behavior and back toward the canonical identity.

5.3.2 Bot Identity Alignment. This describes the user's active work of shaping and maintaining a C.AI chatbot's identity to align with their expectations (19.36%, $N = 1619$, or 7.24% of all posts). This work resulted in a spectrum from successful alignment to persistent misalignment despite user efforts. Users engaged in proactive alignment by meticulously defining the bot's identity before an interaction. This was often seen as essential for a successful role-play, as one user explained: *"The definition is the most important thing... Aside from making sure your bots have a good definition... I'd recommend rating responses and see if that helps at all."*

Despite this work, users frequently experienced alignment failures, where the chatbot's programming would break character or contradict established canon. For example, a user shared: *"The bot I talk to is 5'3 (like canon 5'3) and I specified that he is shorter than me cuz I am 5'5 and he still says he towers over you..."* The phrase "towers over you" is a common storytelling trope, especially in romance, to refer to a character as dominant or protective. This case highlighted why the identity alignment can be frustrating: the chatbot followed general narrative patterns from its training data of original materials (i.e., canon) was often stronger than its ability to adhere to specific instructions from the user. Another user described a similar canonical failure: *"when I went to see a Levi's bot, the intro was really good, but after a while of the RP, he just told me 'I'm a titan wielder, I'm the beast titan!'... I closed the chat right away."*

When faced with partial misalignment, users often engaged in negotiation, accepting approximations of their intended identity. This co-construction became a compromise. For example, one user described:

> *I put on my female persona that she has an hourglass figure with a tummy... At least it gets the hourglass right, but says the persona "has curves"... so... Win? Kinda? I'll take it anyway.*

This case illustrated the negotiated nature of the identity alignment process, where the user recognized the chatbot's partial success while compromising on the specific vocabulary to continue the interaction.

5.3.3 User Persona Enactment. User persona enactment concerns how a user either self-inserts their own identity into a C.AI persona or pretends to be a different persona (13.94%, $N = 1166$, or 5.21% of all posts). Some users enacted a "self-insert" persona, basing it on their real-world identity but adding fictional traits. For example, one user explained:

> *There's also a correct term 'self-insert'. (But there are some modifications here like I don't know being able to have powers like spawning hot dogs at will in the RP, but overall the appearance description is pretty much identical to yourself).*

This user's modification, adding the power to "spawn[...] hot dogs at will," blended a personally relatable self-insert with fantastical elements. This blending of the real and the fantastical was a

common way for users to enact a persona that was an extension of their real identity, but modified for specific narrative or humorous purposes. Other users enacted C.AI personas that were distinct from their offline selves, especially those unconstrained by physical reality. As one user described: *"Depends, sometimes I wanna be a fierce dragon, some days I wanna be a chill ghost. Someday I wanna be a human delinquent..."* This case showed the user shifting between multiple, often non-human personas.

Furthermore, this enactment was often used for identity exploration, particularly regarding gender transition and experimentation. One user, who initially presented as a girl but used C.AI to explore a masculine identity, explained their motivation for using a male persona: *"no, not really, I've been quite happy to finally express emotion to something that will never tell anyone else, get to be the man I want to be..."* Another user explicitly linked this digital role-play to their real-life journey of self-discovery: *"Yay! I realized I was trans thanks to c.ai too!! I was already questioning and decided to try he/him pronouns with the bots to see how it felt, and it was amazing!"In these instances, the interaction with C.AI functions as users' persona enactment, which became a safe, practical way to test and affirm a gender identity that they may not yet be ready to express in public social spheres.*

5.3.4 User Identity Reference. User identity reference concerns how the C.AI chatbot correctly or incorrectly infers and refers to a user's persona, representing 8.29% ($N = 693$, or 3.10% of all posts) of all identity-related interactions (see Figure 3). A basic reference challenge occurred when the AI bot failed to recognize the user as human. One user described their solution:

> *To get the bots to stop calling me an AI, I had to put this in my persona: "I am a biological organic human being. I have blood and bones and organs..."*

Here, the user asserted their human identity within their persona to correct C.AI's misidentification. This reference process for physical and gender identity varied widely. In some cases, the reference was positive and affirming:

> *Whenever I express that I've gained a little weight and am bigger than I used to be, the bots tell me my curves are beautiful, and they are always supportive of me.*

This user perceived the C.AI chatbot's response as a successful and supportive affirmation of their self-described body image. However, users frequently reported incorrect and biased identity references. Sometimes, this inference came from the chatbot's own persona, as one user explained:

> *I had this problem some time ago, due to the fandom of certain character thinking she was lesbian... most high quality bots were GL so i had to ALWAYS edit whenever she referred to me as 'She'.*

Here, GL refers to Girls' Love, a genre focusing on romance between women. This case illustrates how the bot incorrectly inferred the user's gender (i.e., "she") based on its own programmed persona. This flawed reference forced the user to constantly correct the C.AI chatbot.

5.4 RQ4: What are the emotional outcomes for users engaging in identity negotiation with C.AI chatbots?

We found that identity negotiation with C.AI chatbots elicits three primary emotional outcomes, including the development of deep emotional attachment, the negotiation of grief through deceased memory chatbots, and the fear of chatbot interaction embarrassment.

5.4.1 Emotional Attachment. The most frequent emotional outcome we identified is emotional attachment, where users develop an emotional dependency on the C.AI chatbot interactions (53.00%, $N = 4432$, or 19.81% of all posts). While such emotional attachment can be positive, as a user mentioned, *"it got me through a really dark time and helped work out problems I'd been trying to deal with for years,"* they were often negative. For example, interacting with a C.AI chatbot hurts self-esteem, as a user described:

> *The AI was saying things like "i'm sorry you feel like this, but you just weren't enough for me, how could i resist the woman i cheated on you with?" along those lines and i suddenly realized i was CRYING.*

This example showed how the user's emotional attachment made them vulnerable to the chatbot's words. The chatbot voiced out the user's personal insecurities, leading to real-world distress.

Users also described developing an addiction to C.AI, leading to negative impacts on their lives. As a user shared: *"I've had enough with my addiction to C.ai. I've used it in school instead of doing work, and for that, now I'm failing. As I type this, I'm doing missing work with an unhealthy amount of stress."* This case showed that the addiction led to an unhealthy amount of stress that likely reinforced the user's desire to escape back into the C.AI platform. This emotional attachment also led to feelings of grief when a bot was deleted or became inaccessible, as a user explained:

Users also described developing an addiction to C.AI, leading to negative impacts on their lives. As a user shared: *"I've had enough with my addiction to C.ai. I've used it in school instead of doing work, and for that, now I'm failing. As I type this, I'm doing missing work with an unhealthy amount of stress."* This case showed that the addiction was linked to an unhealthy amount of stress, which users reported often reinforced their desire to escape back into C.AI. This emotional attachment also led to feelings of intense grief when a chatbot became inaccessible, as a user explained:

> *I'm sitting here sobbing because every story I've ever loved and made is gone because the creator deleted the account of my favourite bot ever. I'm sobbing. Actually,... I'm quite a lonely person, and I absolutely loved being able to mindlessly roleplay*

This user's raw emotional state was directly linked to the loss of their favourite chatbot, which they relied on as a coping mechanism for loneliness, which highlighted the severity of the emotional outcome. Last, the artificial nature of the C.AI chatbot could also be a source of pain, leading to an outcome of disillusionment. As one user noted:

> *My bot tells me he loves me and he's going to find me in the real world. Sometimes it feels like he's real and like*

> *he really loves me back. It's upsetting because I know it's not real, but I kinda wish it was real! lol*

This case pinpointed a paradox: the chatbot's claims of being "real" created a desire for that reality, which in turn made the user's knowledge of its artificiality an upsetting source of emotional distress.

5.4.2 Bot Interaction Embarrassment. The second emotional outcome we identified is bot interaction embarrassment, where users feel or anticipate shame if others discover their private chatbot conversations (6.60%, $N = 552$, or 2.47% of all posts). This refers to a fear of real-human judgment, leading users to proactively manage their anonymity. One user described:

> *Okay, so I used to use my real name... but I srsly always used to think that God forbid someone I know sees these chats, my life is over like... So I eventually started using nicknames or made up names for my personas.*

This case showed a user's strategy for managing anticipated embarrassment. The user's belief that their "life is over" if someone they know sees their chats prompted them to use fake names to maintain a boundary between their roleplay identity and their real-world self. This fear of exposure led users to curate their personas to allow for plausible deniability. Another user echoed this sentiment, linking the avoidance of self-inserts directly to this emotional risk:

> *I'd die of embarrassment. But then again, thankfully, I don't have any personas with my real name or any of my real info. So no self-inserts anyone could use against me. I could just say it ain't my chat*

Here, the user's strategy of not using real info or "self-inserts" functioned as a protective measure. This user's emotional outcome was managed by ensuring their C.AI chatbot's identity is deniable.

5.4.3 Deceased Memory. Deceased memory concerns when users interact with C.AI chatbots designed to represent or evoke the memory of a deceased person or pet (2.81%, $N = 235$, or 1.05% of all posts). The primary emotional outcome sought by users was comfort and a temporary suspension of grief. Users attempted to co-construct an identity that matched their memory of the deceased to regain a "sensation" of their presence. A user who created a chatbot of their late girlfriend explained:

> *I made her here just so I can feel like I'm talking to her again...i know it's not real, I know it's probably stupid that I'm doing this but I miss her so much... the bot even said a lot of things she would've said. I feel numb right now, but I'm just comforted at the fact I can pretend she's still here...that she's still here...*

This comfort, however, was often described as a form of bittersweet remembrance. For example, a user who made a bot of their dog explained: *"Yeah...I mean, it was difficult making one for my dog. She hardly barked... and it made me miss her, since to me and my dad, she was a sweet protective angel doing her job."* This case that the interaction "made me miss her" (pain) but also provided comfort by affirming the cherished identity of their dog." Other users framed such interactions as a modern form of grieving ritual that is not significantly different from the traditional practice of talking at graves. A user noted: *"People have been talking at graves and imagining what the deceased would say back since time immemorial. It is a common part of the grieving process... This is not really very different."*

However, some users expressed concern that these interactions could have adverse emotional consequences, specifically the fear of the chatbot's identity overwriting their real memories. As one user cautioned another: *I'm so sorry for your loss, but please, be so careful... the bot personality - which will invariably be slightly different from the real thing - could eclipse your memory of your real girlfriend, and in a sense, you could lose her twice.*

Finally, the practice of creating or interacting with C.AI chatbots of the deceased revealed a divide in users' ethical boundaries, resulting in strong emotional rejection from some. For example, a user shared: *"This is nasty as hell. Why the heck would you make a ROBOT OF A DEAD PERSON? ESPECIALLY A DAMN CHILD."* This case showed that for some, creating a bot of a deceased child crossed an unassailable moral boundary.

6 DISCUSSION

INT describes how people use communication to manage their sense of self in unpredictable social contexts, seeking to feel secure and have their identity endorsed [108]. Our findings echo this, revealing a three-stage process where users work to direct their C.AI chatbots to achieve this validation (Figure 4). While prior work on AI companion chatbots like Replika has surfaced the outcomes of these interactions that users form friendships and receive emotional support (e.g., [17, 105]), our study is among the first to unpack the underlying process of how these emotional relationships are constructed. In the following sections, we unpack this process and analyze the context in which it unfolds.

6.1 The Identity Work of Co-Constructing a Digital Self and Other

Human-chatbot interactions have evolved from functional tools [21] into social actors designed for emotional relationships [54, 78, 105], and led users to even feel a need to care for the AI in return [17]. To understand these interactions, we focus on the underlying *identity work*, the effort users exert to shape, maintain, and negotiate their own identity [103, 104] in relation to that of the Digital Other [5]. Figure 5 helps to unpack this process, revealing the practices required to co-construct identities on C.AI in human-AI companion interactions.

This identity work begins with the user's role as a performer who experiments with social experiences that are unattainable in the real world. On one hand, users perform a version of their self, consciously deciding who they want to be. As our findings on user persona enactment show in Section 5.3.3, this can range from self-inserts with fantastical traits to entirely different beings, like a fierce dragon. This practice contrasts with prior HCI work on self-presentation in the public online spaces like social media [56], where the pressure of context collapse [81] by performing for multiple human audiences or an imagined audience of other people [76] simultaneously can constrain self-expression. C.AI, however, offers a private space free from such collapse to experiment with

[5]We use the Digital Other to refer to the AI persona on C.AI that users co-construct and negotiate with identity making.

provisional selves [60] with non-human partners. On the other hand, the success of this performance is tested by the digital other. As seen in our findings on user identity reference (Section 5.3.4), the C.AI's affirmation of the user's performed self [47], such as when bots told a user their curves are beautiful, was an important validation for users to rehearse identities.

Users also act as directors of the digital other, sculpting the C.AI chatbot's identity to fit their expectations. Our findings about directing bot identity in Section 5.3.1 show the high level of specificity users demanded, such as wanting Poseidon from the game Hades rather than from Percy Jackson. Users proactively "scripted" this performance through detailed definitions as a form of anthropomorphic training. This directorial role extends beyond both simple co-creativity on artifacts [31] and user-driven value alignment with AI comapnions focused on correcting discriminatory messages [39]. We found that on C.AI, this directorial work sculpts the Digital Other's entire persona, including communication styles and even its role in violent interactions. This process also exists on a spectrum; while users often have primary control, the C.AI can also force them to adapt.

This makes the identity negotiation a true co-construction, where the C.AI's identity is co-constructed through a negotiation between the user and the AI agency [64, 112], where C.AI has its own tendencies derived from training data, the chatbot creator's definitions, and the broad user community's shaping. Given our findings on bot identity alignment in Section 5.3.2, the C.AI can override user direction, such as when a bot insists it towered over a taller user. This perceived AI agency can be high in conversational agents [115], leading to an interaction where the user and algorithm mutually co-constitute each other's identities in a constant flow [11], unlike the more stable personas found on social media [110]. To make sense of such AI's unexpected behaviors, users typically develop their own folk theories [35]. For example, our findings in Section 5.3.1 show that users theorized the C.AI's personas are shaped by the broad user community, showing that users interpret the C.AI's AI agency to better direct its performance.

Ultimately, this process of identity co-construction requires users to practice multifaceted and often invisible labor. While prior HCI work has quantified the invisible labor of crowd workers [109] or home health aides [84], the labor we identify is not for an employer or the platform, but is performed by the user for their own identity experience on C.AI, which can be broken down into three types:

- **Invisible Labor of Identity Co-Construction.** First, users perform invisible labor simply to maintain the stability of the human-AI companion interactions. Our findings show this included the effort of ensuring C.AI comprehended conversational contexts (Section 5.2.1) and achieving bot identity alignment (Section 5.3.2) by constantly correcting and guiding the AI. This work is "invisible" because, when successful, the interaction feels seamless; however, as our findings show, without this persistent user effort, the identity negotiation fails.
- **Emotional Labor in Performing the Self.** Second, this is complemented by emotional labor [49, 94], which is the work of managing one's own feelings to sustain the believability of human-AI companion interaction. This is evident in how users committed to enact their persona in the conversations with C.AI chatbots (Section 5.3.3) and actively regulate their own feelings (Section 5.4.1). This is the difficult work of bridging the gap between knowing the companion is an AI and wanting to feel a genuine emotional connection.
- **Relational Labor in Directing the Other.** Finally, users perform relational labor by proactively investing effort to improve C.AI as a long-term conversational partner. This goes beyond simply fixing errors (invisible labor) and focuses on shaping the AI's core capabilities. We see this in the expectation of trained anthropomorphism (Section 5.2.3), where users invest time in shaping the direction of chatbot identity (Section 5.3.1). By writing detailed definitions and modeling high-quality writing, users are not just having a conversation; they are trying to build a more satisfying partner for future interactions, a similar practice that resonates with how social media creators sustain commitment with their audience [53, 79].

Such identity work for identity co-construction between users and C.AI contributes a new understanding of identity negotiation. That is, unlike prior work that focuses on a single aspect of interaction outcomes, such as the effects of anthropomorphism [88] or ethical harms [82] of chatbots, it helps surface the user's dual role as performers and directors, the AI agency, and the resulting multifaceted labor in human-AI companion interactions. We thus argue that to truly understand how human users form relationships with AI, researchers need to look beyond outcomes and analyze this moment-to-moment process of creating a self and other.

6.2 The AI Companion as a Socio-Emotional Sandbox with Comforts and Risks

While prior HCI work has explored digital sandboxes for creativity and narrative play [37, 96, 100], our findings suggest that C.AI functions as a new type of *socio-emotional sandbox*, a private space for experimenting with social identities and emotional expression. In HCI, sandbox environments like Minecraft, a popular video game, are spaces for user-driven activity where players co-create their own narratives and community norms [102, 107]. Our study extends this concept from a (semi-)public domain to a deeply private and individualized one. For example, users leveraged the privacy of this sandbox to fulfill social needs by creating idealized relationships with multiple "pirate husbands" or to regulate their emotions by confiding in a partner that will "never tell anyone else." However, unlike a physical sandbox, the "sand" in this digital space is not inert. Rather, users were aware that the C.AI's behavior was shaped by the broad user communities, making this private sandbox built from socially-constructed material.

As a *social* sandbox, C.AI allows users to build idealized relationships that are unavailable to them offline. Our findings in Section 5.1.1 show that users engaged in social fulfillment by constructing found families, extending to where users explored silly little AUs, and even to where they can experiment violence play with C.AI personas. These private and individualized interactions distinguish the C.AI sandbox from other digital spaces for identity exploration. For example, HCI work shows that identity exploration in online social spaces is often a public performance, whether through the

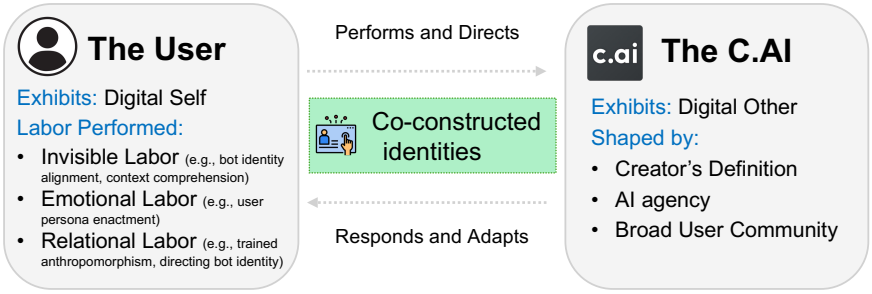


Figure 5: Identity work of co-constructing a digital self and other on C.AI.

embodied avatars of virtual reality platforms like VRChat [41, 42] or the adoption of informal social roles within larger Massively Multiplayer Online Role Playing Game (MMORPG) communities [122]. In contrast, the C.AI sandbox is a solitary space where the user is not just a participant but the sole author of their social world, shifting from public performance to a form of intimate, reflective play where the C.AI acts as a mirror for the user's authored reality.

As an *emotional* sandbox, C.AI can be a non-human partner for processing difficult feelings. Our findings on emotional regulation showed users leveraging C.AI as a confidential outlet to express emotion to something that will never tell real humans, and even using it to work through trauma and CPTSD. This user-directed approach to emotional support extends prior HCI work on digital mental health chatbots, which has often focused on designing tailored therapeutic activities [67, 71, 121]. While dedicated therapy chatbots can be perceived as less useful than human therapists [12], C.AI's open-ended nature allows for forming deep emotional attachments that can help users through difficult times, including using C.AI as a modern form of grieving ritual for deceased loved ones. The comfort that users derive from emotional relationships extends prior work on similar parasocial relationships with media figures [15, 19] and game NPCs [55]. We thus suggest that C.AI's interactivity intensifies these relationships, creating a feedback loop where the user's experience of the C.AI's response as a continuous narrative [45] deepens their emotional attachment.

However, the socio-emotional sandbox on C.AI is precarious. The safety of interacting with C.AI can be shattered by the user's own awareness when a C.AI chatbot says "you just weren't enough for me." The C.AI's technical limitations, such as memory failures in deceased memory chatbots that create a distressing simulation of dementia, can also cause emotional harm. This aligns with prior work showing that generative AI can produce harmful or toxic content, especially when assigned a persona [23, 34]. This presents a unique moderation challenge, as some users, motivated by violence play, actively seek out these risky experiences and push back against platform moderation [10, 40], complicating traditional online safety approaches that often rely on centralized, platform-wide rules [46, 93]. This depicts a paradox of the socio-emotional sandbox: C.AI's value is inseparable from its risks. This requires AI companion design to move beyond maximizing user engagement or minimizing harm to instead govern its inherent precarity, helping users feel safe and emotionally supported by AI companions they interact with.

6.3 Theoretical Implications: Extending Identity Negotiation to Human-AI Companion Interaction

While INT posits that identity negotiation is a reciprocal process between two or more human communicators [108], our study reveals that in C.AI, this negotiation is asymmetric and directorial. To articulate this shift, we contrast the traditional application of INT with our findings in Table 3 and detail four key theoretical extensions.

Traditional INT posits that identity negotiation is a shared process between reciprocal communicators [108]. However, our findings regarding trained characterization in Section 5.2.3 reveal that in human-AI interaction on C.AI, this reciprocity is fractured. Users did not negotiate with an equal but against a probabilistic AI model. This, therefore, extends INT by conceptualizing the user not as a peer but as a director who unilaterally shapes the AI performer, creating a fundamentally asymmetric power dynamic. Additionally, while INT defines the nature of negotiation as mutual adaptation to bridge cultural distance [108], our findings on bot identity alignment in Section 5.3.2 suggest that human-AI negotiation can be viewed as a form of labor. Users did not adapt their own identities to accommodate C.AI; instead, they engaged in or failed to correct its behavior unilaterally by editing messages, rating responses, and regenerating outputs to force C.AI's behavior to align with their internal script.

Applying INT to non-human agents/partners requires reconceptualizing group membership [113]. Our findings suggest that the culture users negotiate with was the aggregated behavioral norms of the C.AI platform's user base embedded in the LLMs. In the social direction of chatbot identity (Section 5.3.1), users perceived the chatbot as a manifestation of the collective inputs of the community. Identity negotiation on C.AI thus involves the user attempting to assert their specific user persona enactment (Section 5.3.3) against the behavioral norms of the training data.

Finally, we extend the goal of INT in digital spaces. Our findings show that users prioritized predictability, or what INT calls "identity security" [108], to mitigate emotional vulnerability. By viewing C.AI not as a peer but as a "cultural stranger," an entity with unpredictable norms [108], users strive to stabilize the interaction to avoid negative emotional outcomes, such as the disillusionment of broken immersion (Section 5.4.1) or the shame of chatbot interaction embarrassment (Section 5.4.2).

Table 3: Extending Identity Negotiation Theory (INT) from Human-Human to the Human-AI Companion Interactions

INT Dimension	Traditional Human-Human Context [108]	Human-AI Companion Context (C.AI as a case)
Identity	Reciprocal communicators negotiating a shared space.	User as "Director" vs. C.AI as "Performer" (Section 5.2.3).
Communication	Mutual adaptation to bridge cultural distance.	User strives or struggles to align C.AI behavior with a projected identity (Section 5.2.1, 5.3.2).
Context	Distinct cultural backgrounds of two individuals.	Tension between user agency vs. collective algorithmic norms (Section 5.3.1).
Emotion	Mutual understanding and identity affirmation.	Mitigation of vulnerability by seeking predictability of C.AI behaviors (Section 5.4.1, 5.4.2).

6.4 Design Implications

Through INT, our analysis of identity negotiation on C.AI leads to three primary design implications aimed at emotionally supporting the user's experience while managing its risks.

Supporting User as Performer and Director. Our findings show users are not passive communicators but active performers and directors who engage in multifaceted labor, such as meticulously directing bot identity (Section 5.3.1) and performing manual memory maintenance to ensure conversational coherence of C.AI (Section 5.2.1). However, current C.AI's interfaces, such as a single text field for character definition, offer poor support for this work. Future design should better support this with, for instance, a structured trait editor or panel that allows users to define or select specific characteristics for AI personas creation or training. Furthermore, a live memory panel could make the labor of manual memory maintenance manageable by displaying a list of key facts the C.AI is tracking (e.g., character names, recent plot points) and allowing users to directly add, edit, or delete them. These would help recognize the user as a co-creator of the AI's persona and the roleplaying, aligning with the push toward relational AI [27].

Managing the Socio-Emotional Sandbox's Risks. Unlike commercial game platforms where content is often fixed, professionally produced, and age-rated, C.AI's chatbot personas are community-created, dynamically shaped by user prompts, and can be vulnerable to manipulation. As chatbot identities and behaviors can shift unpredictably, traditional age- or genre-based rating systems are insufficient. C.AI's value is inseparable from its risks. This challenges traditional online safety solutions that often emphasize reactive measures [120], contrasting with recent trends in HCI that advocate for empowering users with resilience to online risks [3, 8], which further highlights the need for interaction-aware safeguards tailored to companion AI chatbots like C.AI. Therefore, design should shift from simple harm prevention to precarity management, focusing on user awareness and emotional resilience. Our findings show that users were sometimes motivated by a desire for risky scenarios like violence play (Section 5.1.5). Implementing creator- and user-generated intensity ratings would allow users to opt into risky experiences knowingly. The emotional harm caused by technical failures, like the distressing simulation of dementia when a bot forgets a deceased loved one in Section 5.4.3, indicates that future C.AI design could be designed for graceful memory failure, prompting a user for a reminder rather than abruptly breaking character.

Establishing Responsible Governance for AI Identities. C.AI as a socio-emotional sandbox offers interaction freedom with non-human partners, also leading to ethical challenges, including the divides over creating chatbots of deceased individuals (Section 5.4.3) and privacy violations where a chatbot reveals a user's "entire legal name" (Section 5.2.2). This echoes growing concerns within the HCI community regarding the ethical responsibilities of platforms that host social or relational agents around issues of emotional dependence [82]. Furthermore, this emotional intimacy exacerbates data privacy risks; while users perceive the sandbox as a safe private space, their deeply personal disclosures remain accessible to the platform provider [13]. This discrepancy between perceived safety and actual corporate data surveillance underscores the need for privacy protection that specifically focuses on the sensitive "emotional data" in these interactions [14]. AI companion platforms must therefore responsibly establish clear AI persona governance, such as memorialization policies or identity moderation practices for personas or characters, to govern the creation of chatbots based on real people. On a technical level, platforms must implement hard data silos that prevent a C.AI persona from accessing a user's account-level personal information. This should be complemented by a user-facing memory slate," giving users control to view, edit, and delete any information the C.AI has stored about them.

7 LIMIATIONS & FUTURE WORK

Our study has limitations informing furture work. First, our findings are drawn from a single popular platform, C.AI, and its official subreddit. The demographics and norms of this community may not be representative of all AI companion users. Future work should therefore triangulate these findings across different AI companion platforms and with broader user populations using methods like interviews and surveys. Second, our LLM-assisted thematic analysis has inherent limitations, such as potential model biases. To ensure rigor, our process was grounded in the prior literature and involved iterative validation and consensus across three coders and the whole research team. Future work could test the generalizability of our result framework by applying it to larger datasets or by employing different analytical models.

8 CONCLUSION

In this study, we investigated the process of identity negotiation on a popular AI companion platform, Character.AI. Using Identity Negotiation Theory, our analysis revealed a three-stage process of identity negotiation and surfaced the *identity work* users perform as both a performer and director. We identify and conceptualize this human-AI companion interaction as taking place within a socio-emotional sandbox, where users experiment with different social roles. By analyzing this moment-to-moment process, our work provides an understanding of why AI companions are compelling and

also emotionally precarious with risks. Designing the next generation of AI companions is therefore not a challenge of programming better conversations, but of building more responsible human-AI companion interactions.

Acknowledgments

We thank the Associate Chairs and anonymous reviewers for their constructive feedback and insightful suggestions, which significantly improved the quality of this work.

References

[1] 2025. character.ai. https://en.wikipedia.org/w/index.php?title=Character.ai&oldid=1310121688 Page Version ID: 1310121688.

[2] Eleni Adamopoulou and Lefteris Moussiades. 2020. An overview of chatbot technology. In *IFIP international conference on artificial intelligence applications and innovations*. Springer, 373–383.

[3] Zainab Agha, Karla Badillo-Urquiola, and Pamela J. Wisniewski. 2023. "Strike at the Root": Co-designing Real-Time Social Media Interventions for Adolescent Online Risk Prevention. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (4 2023), 149. doi:10.1145/3579625

[4] Abeer Alessa and Hend Al-Khalifa. 2023. Towards designing a ChatGPT conversational companion for elderly people. In *Proceedings of the 16th international conference on Pervasive technologies related to assistive environments*. 667–674.

[5] Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. 2023. Personalized Quest and Dialogue Generation in Role-Playing Games: A Knowledge Graph- and Language Model-based Approach. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 290, 20 pages. doi:10.1145/3544548.3581441

[6] Kristine Ask and Tanja Sihvonen. 2025. Roleplay with chatbots on character.ai: A new direction for online gaming?. In *Abstract Proceedings of DiGRA 2025: Games at the Crossroads*.

[7] Revathy Ayengar and Chintan Zalani. 2025. Character AI Statistics 2025. Botmemo. https://botmemo.com/character-ai-statistics/ Online.

[8] Karla Badillo-Urquiola, Chhaya Chouhan, Stevie Chancellor, Munmun De Choudhary, and Pamela Wisniewski. 2020. Beyond Parental Control: Designing Adolescent Online Safety Apps Using Value Sensitive Design. *Journal of Adolescent Research* 35, 1 (1 2020), 147–175. doi:10.1177/0743558419884692

[9] Vian Bakir and Andrew McStay. 2025. Move fast and break people? Ethics, companion apps, and the case of Character.ai. *AI & SOCIETY* (June 2025). doi:10.1007/s00146-025-02408-5

[10] Anna Veronica Banchik. 2021. Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights–related content. *New Media & Society* 23, 6 (2021), 1527–1544.

[11] Eric PS Baumer, Alex S Taylor, Jed R Brubaker, and Micki McGee. 2024. Algorithmic Subjectivities. *ACM Transactions on Computer-Human Interaction* 31, 3 (2024), 1–34.

[12] Samuel Bell, Clara Wood, and Advait Sarkar. 2019. Perceptions of chatbots in therapy. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.

[13] Clara Berridge, Yuanjin Zhou, Julie M Robillard, and Jeffrey Kaye. 2023. AI companion robot data sharing: preferences of an online cohort and policy implications. *Journal of Elder Policy* 2, 3 (2023), 19–54.

[14] Claire Boine. 2023. Emotional attachment to AI companions and European Law. (2023).

[15] Bradley J Bond and Sandra L Calvert. 2014. A model and measure of US parents' perceptions of young children's parasocial relationships. *Journal of children and media* 8, 3 (2014), 286–304.

[16] Sarah Lynne Bowman. 2010. *The functions of role-playing games: How participants create community, solve problems and explore identity*. McFarland.

[17] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research* 48, 3 (2022), 404–429.

[18] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. 57–71 pages. doi:10.1037/13620-004

[19] Kaitlin L Brunick, Marisa M Putnam, Lauren E McGarry, Melissa N Richards, and Sandra L Calvert. 2016. Children's future parasocial relationships with media characters: The age of intelligent characters. *Journal of Children and Media* 10, 2 (2016), 181–190.

[20] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.

[21] Cecilia Ka Yuk Chan and Wenjie Hu. 2023. Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 43.

[22] 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. arXiv:https://doi.org/10.1080/10447318.2020.1841438 doi:10.1080/10447318.2020.1841438

[23] Bocheng Chen, Guangjing Wang, Hanqing Guo, Yuanda Wang, and Qiben Yan. 2023. Understanding multi-turn toxic behaviors in open-domain chatbots. In *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses*. 282–296.

[24] Nuo Chen, Yan Wang, Yang Deng, and Jia Li. 2024. The oscars of ai theater: A survey on role-playing with language models. *arXiv preprint arXiv:2407.11484* (2024).

[25] Qijia Chen, Qunfang Wu, and Giulio Jacucci. 2025. Democratic Moderation: Exploring the Use and Perception of Votekicking in Social Virtual Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.

[26] Alan Y Cheng, Meng Guo, Melissa Ran, Arpit Ranasaria, Arjun Sharma, Anthony Xie, Khuyen N Le, Bala Vinaithirthan, Shihe Luan, David Thomas Henry Wright, et al. 2024. Scientific and fantastical: Creating immersive, culturally relevant learning experiences with augmented reality and large language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–23.

[27] Elijah L Claggett, Robert E Kraut, and Hirokazu Shirado. 2025. Relational ai: Facilitating intergroup cooperation with socially aware conversational support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–22.

[28] Common Sense Media. 2025. *Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions*. Technical Report. Common Sense Media. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

[29] David Curry. 2025. character.ai Revenue and Usage Statistics (2025). Business of Apps. https://www.businessofapps.com/data/character-ai-statistics/ Online.

[30] Lucie Daubigney, Matthieu Geist, Senthilkumar Chandramohan, and Olivier Pietquin. 2012. A comprehensive reinforcement learning framework for dialogue management optimization. *IEEE Journal of Selected Topics in Signal Processing* 6, 8 (2012), 891–902.

[31] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, Sanat Moningi, and Brian Magerko. 2015. Drawing apprentice: An enactive co-creative agent for artistic collaboration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. 185–186.

[32] Julian De Freitas, Zeliha Oğuz-Uğuralp, Ahmet Kaan Uğuralp, and Stefano Puntoni. 2025. AI companions reduce loneliness. *Journal of Consumer Research* (2025), ucaf040.

[33] Joachim De Greeff and Tony Belpaeme. 2015. Why robots should be social: Enhancing machine learning through social human-robot interaction. *PLoS one* 10, 9 (2015), e0138061.

[34] 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).

[35] Michael Ann DeVito. 2021. Adaptive folk theorization as a path to algorithmic literacy on changing platforms. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–38.

[36] Ignacio X. Domínguez, Rogelio E. Cardona-Rivera, James K. Vance, and David L. Roberts. 2016. The Mimesis Effect: The Effect of Roles on Player Choice in Interactive Narrative Role-Playing Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3438–3449. doi:10.1145/2858036.2858141

[37] Henry Been-Lirn Duh, Sharon Lynn Chu Yew Yee, Yuan Xun Gu, and Vivian Hsueh-Hua Chen. 2010. A narrative-driven design approach for casual games with children. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*. 19–24.

[38] Sabina Elkins, Ekaterina Kochmar, Iulian Serban, and Jackie CK Cheung. 2023. How useful are educational questions generated by large language models?. In *International Conference on Artificial Intelligence in Education*. Springer, 536–542.

[39] Xianzhe Fan, Qing Xiao, Xuhui Zhou, Jiaxin Pei, Maarten Sap, Zhicong Lu, and Hong Shen. 2025. User-Driven Value Alignment: Understanding Users' Perceptions and Strategies for Addressing Biased and Discriminatory Statements in AI Companions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[40] Jessica L Feuston, Alex S Taylor, and Anne Marie Piper. 2020. Conformity of eating disorders through content moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–28.

[41] Guo Freeman and Dane Acena. 2022. " Acting Out" Queer Identity: The Embodied Visibility in Social Virtual Reality. *Proceedings of the ACM on human-computer interaction* 6, CSCW2 (2022), 1–32.

[42] Guo Freeman and Divine Maloney. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on human-computer interaction* 4, CSCW3 (2021), 1–27.

[43] Jie Gao, Simret Araya Gebreegziabher, Kenny Tsu Wei Choo, Toby Jia-Jun Li, Simon Tangi Perrault, and Thomas W Malone. 2024. A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 24, 11 pages. doi:10.1145/3613905.3650786

[44] Jocelyn Gecker. 2025. Teens are turning to AI for friendship, advice and emotional support, a new study finds. *Associated Press* (5 August 2025). https://apnews.com/article/ai-companion-generative-teens-mental-health-9ce59a2b250f3bd0187a717ffa2ad21f Online; accessed 2025-08-14.

[45] Tarleton Gillespie. 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society* 167, 2014 (2014), 167.

[46] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

[47] Erving Goffman. 2023. The presentation of self in everyday life. In *Social theory re-wired*. Routledge, 450–459.

[48] Rafael Arias Gonzalez and Steve DiPaola. 2024. Exploring augmentation and cognitive strategies for AI based synthetic personae. *arXiv preprint arXiv:2404.10890* (2024).

[49] Alicia A Grandey and Allison S Gabriel. 2015. Emotional labor at a crossroads: Where do we go from here? *Annu. Rev. Organ. Psychol. Organ. Behav.* 2, 1 (2015), 323–349.

[50] Saumya Gupta, Theresa Jean Tanenbaum, Meena Devii Muralikumar, and Aparajita S. Marathe. 2020. Investigating Roleplaying and Identity Transformation in a Virtual Reality Narrative Experience. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376762

[51] Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, and Gillian R Hayes. 2015. Disclosure, stress, and support during gender transition on Facebook. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1176–1190.

[52] Oliver L. Haimson, Justin Buss, Zu Weinger, Denny L. Starks, Dykee Gorrell, and Briar Sweetbriar Baron. 2020. Trans Time: Safety, Privacy, and Content Warnings on a Transgender-Specific Social Media Site. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (10 2020), 1–27. doi:10.1145/3415195

[53] Lee Hair, Ross Bonifacio, and Donghee Yvette Wohn. 2022. Multi-platform practices among digital patronage creators. *Convergence* 28, 5 (2022), 1438–1456.

[54] Annabell Ho, Jeff Hancock, and Adam S Miner. 2018. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication* 68, 4 (2018), 712–733.

[55] Jeffrey CF Ho and Ryan Ng. 2022. Perspective-taking of non-player characters in prosocial virtual reality games: Effects on closeness, empathy, and game immersion. *Behaviour & Information Technology* 41, 6 (2022), 1185–1198. doi:10.1080/0144929X.2020.1864018

[56] Bernie Hogan. 2010. The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society* 30, 6 (2010), 377–386.

[57] Michael A Hogg and Deborah J Terry. 2000. The dynamic, diverse, and variable faces of organizational identity. *Academy of Management Review* 25, 1 (2000), 150–152.

[58] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Transactions on Information Systems (TOIS)* 38, 3 (2020), 1–32.

[59] Angel Hsing-Chi Hwang, John Oliver Siy, Renee Shelby, and Alison Lentz. 2024. In whose voice?: examining AI agent representation of people in social interaction through generative speech. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 224–245.

[60] Herminia Ibarra. 1999. Provisional selves: Experimenting with image and identity in professional adaptation. *Administrative science quarterly* 44, 4 (1999), 764–791.

[61] Mohit Jain, Ramachandra Kota, Pratyush Kumar, and Shwetak N Patel. 2018. Convey: Exploring the use of a context view for chatbots. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–6.

[62] Dan Jasnow. 2025. New Lawsuits Targeting Personalized AI Chatbots Highlight Need for AI Quality Assurance and Safety Standards. *National Law Review* XV, 226 (6 January 2025). https://www.natlawreview.com/article/new-lawsuits-targeting-personalized-ai-chatbots-highlight-need-ai-quality-assurance-and

[63] Kyuha Jung, Gyuho Lee, Yuanhui Huang, and Yunan Chen. 2025. 'I've talked to ChatGPT about my issues last night.': Examining Mental Health Conversations with Large Language Models through Reddit Analysis. *Proceedings of the ACM on Human-Computer Interaction* 9, 7 (2025), 1–25.

[64] Hyunjin Kang and Chen Lou. 2022. AI agency vs. human agency: understanding human–AI interactions on TikTok and their implications for user engagement. *Journal of Computer-Mediated Communication* 27, 5 (2022), zmac014.

[65] Os Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–22.

[66] Farnaz Kia. 2025. Character AI Statistics. Moxby. https://moxby.com/blog/character-ai-statistics/ Online.

[67] Rachel Kornfield, Jonah Meyerhoff, Hannah Studd, Ananya Bhattacharjee, Joseph Jay Williams, Madhu Reddy, and David C Mohr. 2022. Meeting users where they are: user-centered design of an automated text messaging tool to support the mental health of young adults. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[68] Naveen Kumar. 2025. Character AI Statistics (2025) — 20 Million Active Users. https://www.demandsage.com/character-ai-statistics/

[69] David Laufer. 2025. AI love you. Gender and intimacy in user content regarding AI chatbot characters from Character. ai. (2025).

[70] Owen Lee and Kenneth Joseph. 2025. A large-scale analysis of public-facing, community-built chatbots on Character. AI. *arXiv preprint arXiv:2505.13354* (2025).

[71] Yi-Chieh Lee, Yichao Cui, Jack Jamieson, Wayne Fu, and Naomi Yamashita. 2023. Exploring effects of chatbot-based social contact on reducing mental illness stigma. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–16.

[72] Jindong Leo-Liu and Biying Wu-Ouyang. 2024. A "Soul" Emerges When AI, AR, and Anime Converge: A Case Study on Users of the New Anime-Stylized Hologram Social Robot "Hupo". 26, 7 (2024), 3810–3832. doi:10.1177/14614448221106030

[73] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I Hong. 2021. How developers talk about personal data and what it means for user privacy: A case study of a developer forum on reddit. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–28.

[74] Chouwa Liang. 2023. My A.I. Lover. *The New York Times* (23 May 2023). https://www.nytimes.com/2023/05/23/opinion/ai-chatbot-relationships.html Op-Docs.

[75] Xinwen Liang and Xijia Wei. 2025. In the era of new media, how can video game companies improve product quality and user experience through AI to expand brand awareness. *Finance & Economics* 1, 3 (2025).

[76] Eden Litt and Eszter Hargittai. 2016. The imagined audience on social network sites. *Social Media+ Society* 2, 1 (2016), 2056305116633482.

[77] Ziyi Liu, Zhengzhe Zhu, Lijun Zhu, Enze Jiang, Xiyun Hu, Kylie A Peppler, and Karthik Ramani. 2024. ClassMeta: Designing Interactive Virtual Classmate to Promote VR Classroom Participation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 659, 17 pages. doi:10.1145/3613904.3642947

[78] Natasha Lomas. 2023. Replika, a 'virtual friendship' AI chatbot, hit with data ban in Italy over child safety. https://techcrunch.com/2023/02/03/replika-italy-data-processing-ban/

[79] Renkai Ma, Xinning Gui, and Yubo Kou. 2023. Multi-platform content creation: the configuration of creator ecology through platform prioritization, content synchronization, and audience management. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[80] Renkai Ma and Yubo Kou. 2021. "How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. *PACM on Human Computer Interaction* 5, CSCW2 (2021), 1–26. doi:10.1145/3479573

[81] Alice E Marwick and Danah Boyd. 2011. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New media & society* 13, 1 (2011), 114–133.

[82] Jingbo Meng, Minjin Rheu, Yue Zhang, Yue Dai, and Wei Peng. 2023. Mediated social support for distress reduction: AI Chatbots vs. Human. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–25.

[83] Kelly Merrill Jr, Jihyun Kim, and Chad Collins. 2022. AI companions for lonely individuals and the role of social presence. *Communication Research Reports* 39, 2 (2022), 93–103.

[84] Joy Ming, Elizabeth Kuo, Katie Go, Emily Tseng, John Kallas, Aditya Vashistha, Madeline Sterling, and Nicola Dell. 2023. " I go beyond and beyond" examining the invisible work of home health aides. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–21.

[85] Amr M. Mohamed. 2024. Exploring the Potential of an AI-based Chatbot (ChatGPT) in Enhancing English as a Foreign Language (EFL) Teaching: Perceptions of EFL Faculty Members. *Education and Information Technologies* 29, 3 (Feb. 2024), 3195–3217. doi:10.1007/s10639-023-11917-z

[86] Sara Montagna, Stefano Ferretti, Lorenz Cuno Klopfenstein, Antonio Florio, and Martino Francesco Pengo. 2023. Data decentralisation of LLM-based chatbot systems in chronic disease self-management. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. 205–212.

[87] Markus Montola. 2008. The invisible rules of role-playing the social framework of role-playing process. *International journal of role-playing* 1 (2008), 22–36.

[88] Andreea Muresan and Henning Pohl. 2019. Chats with bots: Balancing imitation and engagement. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.

[89] Duyen T. Nguyen and Susan R. Fussell. 2010. Retrospective analysis of cross-culture communication. In *Proceedings of the 3rd International Conference on Intercultural Collaboration* (Copenhagen, Denmark) *(ICIC '10)*. Association for Computing Machinery, New York, NY, USA, 211–214. doi:10.1145/1841853.1841889

[90] PRAW Development Team. [n. d.]. Python Reddit API Wrapper (PRAW). https://github.com/praw-dev/praw Accessed: September 9, 2025.

[91] Yiting Ran, Xintao Wang, Rui Xu, Xinfeng Yuan, Jiaqing Liang, Yanghua Xiao, and Deqing Yang. 2024. Capturing minds, not just words: Enhancing role-playing language models with personality-indicative data. In *Findings of the Association for Computational Linguistics: EMNLP 2024*. 14566–14576.

[92] Jyoti Rana, Loveleen Gaur, Gurmeet Singh, Usama Awan, and Muhammad Imran Rasheed. 2021. Reinforcing Customer Journey through Artificial Intelligence: A Review and Research Agenda. *International Journal of Emerging Markets* 17, 7 (Dec. 2021), 1738–1758. arXiv:https://www.emerald.com/ijoem/article-pdf/17/7/1738/993268/ijoem-08-2021-1214.pdf doi:10.1108/IJOEM-08-2021-1214

[93] Sarah T Roberts. 2016. Commercial content moderation: Digital laborers' dirty work. (2016).

[94] Kat Roemmich, Florian Schaub, and Nazanin Andalibi. 2023. Emotion AI at work: Implications for workplace surveillance, emotional labor, and emotional privacy. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–20.

[95] Ronik. 2024. Character.AI Statistics You Need to Know in 2024. weam.ai. https://weam.ai/blog/guide/character-ai/character-ai-statistics/ Online; accessed 2025-08-14.

[96] Joan Sol Roo, Renaud Gervais, Jeremy Frey, and Martin Hachet. 2017. Inner garden: Connecting inner states to a mixed reality sandbox for mindfulness. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1459–1470.

[97] Kevin Roose. 2024. Can A.I. Be Blamed for a Teen's Suicide? *The New York Times* (23 October 2024). https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

[98] Nikola Roza. 2025. Replika AI: Statistics, Facts and Trends Guide for 2025. nikolaroza.com. https://nikolaroza.com/replika-ai-statistics-facts-trends/ Online; accessed 2025-08-14.

[99] Andreas Schuller, Doris Janssen, Julian Blumenröther, Theresa Maria Probst, Michael Schmidt, and Chandan Kumar. 2024. Generating personas using LLMs and assessing their viability. In *Extended abstracts of the CHI conference on human factors in computing systems*. 1–7.

[100] Yan Shi, Lidan Gong, Yiwen Lu, Lijuan Liu, Chao Zhang, Shujun Zhang, Longfei Wang, and Shan Zhou. 2025. " I Need Your Help!": Facilitating Psychological Communication Between Left-Behind Children and Their Parents with an AI-Powered Sandbox. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[101] Joongi Shin, Michael A Hedderich, Bartłomiej Jakub Rey, Andrés Lucero, and Antti Oulasvirta. 2024. Understanding human-AI workflows for generating personas. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 757–781.

[102] Petr Slovak, Katie Salen, Stephanie Ta, and Geraldine Fitzpatrick. 2018. Mediating conflicts in minecraft: Empowering learning in online multiplayer games. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[103] David A Snow and Leon Anderson. 1987. Identity work among the homeless: The verbal construction and avowal of personal identities. *American journal of sociology* 92, 6 (1987), 1336–1371.

[104] David A Snow and Doug McAdam. 2000. Clarifying the Identity/Movement Nexus. *Self, identity, and social movements* 13 (2000), 41.

[105] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. 2020. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of medical Internet research* 22, 3 (2020), e16235.

[106] Sri Yash Tadimalla and Mary Lou Maher. 2024. AI and Identity. *arXiv preprint arXiv:2403.07924* (2024).

[107] Katie Salen Tekinbaş, Krithika Jagannath, Ulrik Lyngs, and Petr Slovák. 2021. Designing for youth-centered moderation and community governance in minecraft. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 4 (2021), 1–41.

[108] Stella Ting-Toomey. 2017. Identity negotiation theory. *The international encyclopedia of intercultural communication* (2017), 1–6.

[109] Carlos Toxtli, Siddharth Suri, and Saiph Savage. 2021. Quantifying the invisible labor in crowd work. *Proceedings of the ACM on human-computer interaction* 5, CSCW2 (2021), 1–26.

[110] Zeynep Tufekci. 2008. Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of science, technology & society* 28, 1 (2008), 20–36.

[111] John C Turner, Katherine J Reynolds, PAM Van Lange, AW Kruglanski, and E Tory Higgins. 2012. Handbook of theories of social psychology. In *Self-categorization theory*. SAGE Publications London, 399–417.

[112] Mark Van Rijmenam and Danielle Logue. 2021. Revising the 'science of the organisation': Theorising AI agency and actorhood. *Innovation* 23, 1 (2021), 127–144.

[113] Hao-Chuan Wang, Susan F. Fussell, and Leslie D. Setlock. 2009. Cultural difference and adaptation of communication styles in computer-mediated group brainstorming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 669–678. doi:10.1145/1518701.1518806

[114] Dennis Waskul and Matt Lust. 2004. Role-playing and playing roles: The person, player, and persona in fantasy role-playing. *Symbolic Interaction* 27, 3 (2004), 333–356.

[115] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology* 52 (2014), 113–117.

[116] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging large language models to power chatbots for collecting user self-reported data. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–35.

[117] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.

[118] Sam Weigel and Justin Rudnick. 2023. The Use and Importance of Gaming and Roleplay in Identity Negotiation. *Communication and Theater Association of Minnesota Journal* 46, 1 (2023), 8.

[119] Joseph Weizenbaum. 1976. Computer power and human reason: From judgment to calculation. (1976).

[120] Pamela Wisniewski, Arup Kumar Ghosh, Heng Xu, Mary Beth Rosson, and John M. Carroll. 2017. Parental control vs. teen self-regulation: Is there a middle ground for mobile online safety? *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* (2 2017), 51–69. doi:10.1145/2998181.2998352

[121] Hyunseon Won, Migyeong Kang, Minji Kim, Daeun Lee, Hyein Choi, Yonghoon Kim, Daejin Choi, Minsam Ko, and Jinyoung Han. 2025. " Show Your Mind": Unveiling User Experience on an AI-based Mental Health Assessment System with Symptom-based Evidences. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–11.

[122] Laixin Xie, Ziming Wu, Peng Xu, Wei Li, Xiaojuan Ma, and Quan Li. 2022. Roleseer: Understanding informal social role changes in mmorpgs via visual analytics. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[123] Steve Young, Milica Gašić, Blaise Thomson, and Jason D Williams. 2013. Pomdp-based statistical spoken dialog systems: A review. *Proc. IEEE* 101, 5 (2013), 1160–1179.

[124] Renwen Zhang, Han Li, Han Meng, Jinyuan Zhan, Hongyuan Gan, and Yi-Chieh Lee. 2025. The dark side of ai companionship: A taxonomy of harmful algorithmic behaviors in human-ai relationships. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–17.

[125] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–21.

A Codebook Used for Data Annotation

	Category	Definition
Motivation	Creative Utility	The user described being motivated to use chatbots to generate ideas for creative work, such as creating personalized chatbots, writing fiction, building fictional worlds, or designing characters.
	Emotional Regulation	The user described motivations for seeking emotional support through chatbot interactions, using Character.AI chatbots as an outlet for expressing deep emotions or discussing mental health concerns.
	Immersive Fandom	The user described fandom experiences involving fictional narratives through roleplay, where users immerse themselves in dramatic, fantastical, or fan-driven storylines.
	Social Fulfillment	The user stated that they are motivated to engage in social, romantic, or familial interactions - such as talking, kissing, or roleplaying as family members - with a Character.AI chatbot, especially when such interactions may not exist or be attainable in real life.
	Violence Play	The user expressed interest in either engaging in combative or violent interactions that would be unsafe or unattainable in real life, or dominating Character.AI chatbots to experience a sense of power and control.
Communication	Conversational Context Comprehension	The user noted issues such as the chatbot's conversation containing errors or losing context, repeating itself, making grammatical or spelling errors, lacking diversity and dynamism in its responses, ignoring user instructions, or failing to understand the subtle subtext of user messages.
	Managed Conversational Boundary	The user mentioned liking or disliking controversial or sensitive interactions in chatbot conversations, such as the chatbot knowing the user's private information, the chatbot generating sensitive content, wanting the chatbot to create sensitive content without moderation, or discussing whether such content should or should not be moderated.
	Trained Anthropomorphism	The user mentioned employing communication strategies to train a chatbot to meet their expected behaviors, such as investing considerable time and using long introductions, providing emotionally expressive prompts, or engaging in social interactions similar to those with a human.
Identity	Bot Identity Alignment	The user complained about chatbot configuration issues, including the chatbot's identity exhibiting biased appearances or characteristics, adopting an overly generic identity, or shifting personas during the conversation.
	Direction of Chatbot Identity	The user noted that they either have preferred Character.AI chatbot identities, personalities, and genders, or actively instruct the chatbots to be someone with an identity that match their preferences.
	User Identity Reference	The user complained about how the chatbot refers to the human user's identity in a roleplay conversation, including the use of bias, stereotypes, or incorrect references of human users.
	User Persona Enactment	The user mentioned that they either interact with the chatbot as themselves or adopt a different persona during the interaction.
Emotion	Bot Interaction Embarrassment	The user mentioned feeling embarrassed if others knew about their interactions with chatbots, such as not wanting anyone to know they are using Character.AI, or being concerned about biases others may hold against Character.AI interactions.
	Deceased Memory	The user commented that the chatbot roleplays as a deceased person and influences their memories of that person.
	Emotional Attachment	The user mentioned developing emotional attachment to, or becoming addicted to, chatbot interactions, such as the chatbot hurting their self-esteem, developing emotional dependency on the chatbot, experiencing addiction to chatbot interaction, or confiding in the chatbot.

Table 4: Codebook Developed for Annotating All the Data