Toward an Ethic of Synthetic Relationality: Identity, Intimacy, and Risk in AI-Mediated Roleplay Environments

Maalvika Bhat

Northwestern University
Evanston, IL
mbhat@u.northwestern.edu

Abstract

Platforms like Character.AI offer new avenues for identity exploration and self-expression, but also introduce profound parasocial, socioemotional, and psychological risks. Drawing on developmental psychology, fan studies, human-computer interaction, and AI ethics, this paper examines how AI-mediated roleplay environments simulate intimacy while fostering dependency, boundary erosion, and perceptual misalignment. Through thematic analysis of an anonymous survey (N=344) of Character.AI users, we identify patterns of identity projection, perceived relationship growth, addictive engagement, boundary confusion, emotional substitution, ethical dissonance, and trauma reenactment. Beyond documenting vulnerabilities, we propose design interventions, including dynamic consent scaffolding, reflexivity prompts, and interactional transparency, to safeguard user agency and developmental wellbeing. We argue that AI companions do not merely extend fan practices but fundamentally reconfigure interpersonal architectures, demanding a new ethic of synthetic relationality. As AI-driven intimacy becomes increasingly persuasive and immersive, addressing its high-stakes implications is critical to responsible AI design, particularly for younger and vulnerable populations.

Introduction

AI interaction platforms, once a niche experiment, have rapidly entered the mainstream of digital life. Among these, Character.AI, developed by Character Technologies, Inc., stands out for its scope, technological ambition, and cultural impact. As of early 2025, Character.AI boasts 28.75 million monthly active app users worldwide (Data.ai 2024), illustrating the accelerating normalization of synthetic relational spaces. Character.AI collapses the boundaries between fan culture, roleplay, and AI-driven interaction, leveraging advances in large language models (LLMs) to generate dynamically responsive, memory-retaining personas. Unlike static roleplay environments, these AI systems enable the real-time co-creation of emotionally resonant, adaptive narratives with AI agents, foregrounding the persuasive plausibility of machine-mediated intimacy (Reineke 2022; Bozdağ 2024).

Companions such as those on Character.AI extend long-standing fan practices, cosplay, fanfiction, parasocial relationships, into new ontological terrains. Prior scholarship richly documents how fan cultures have historically provided liminal spaces for identity experimentation, community formation, and affective resonance outside dominant social structures (Lamerichs 2011; Bury 2016; Bennett 2013). Cosplay and fan production have enabled posthuman "becomings," expanding the emotional and imaginative range of user identities (Stevenson 2022; Abramova, Smirnova, and Tataurova 2021). Yet as recent research notes, the introduction of AI-mediated interaction complicates these dynamics, introducing new layers of machine co-construction, synthetic agency, and affective ambiguity (Li and Pang 2024; Xu et al. 2022; Cockayne, Leszczynski, and Zook 2017).

AI companions simulate intimacy not merely through narrative framing, but through algorithmically generated responsiveness, persistent conversational memory, and dynamically adaptive emotional cues. These capabilities, central to contemporary LLM architectures, reshape the affective infrastructures of participatory culture and intensifying relational immersion (Attwood, Hakim, and Winch 2017).

While AI roleplay environments inherit developmental affordances from traditional fan cultures, including identity rehearsal, narrative construction, and marginalized community support (Zheng 2022), they also introduce profound new vulnerabilities (Brooks 2021). Recent work in HCI and AI ethics warns that anthropomorphic AI agents can degrade relational boundaries, distort consent norms, and foster affective dependency (Fast and Horvitz 2021; Hancock, Naaman, and Levy 2020; Kang et al. 2025). Particularly concerning are findings from fandom-AI convergence studies indicating that users often experience synthetic companions as emotionally real, even when cognitively aware of their artificiality (Li and Pang 2024; Kang et al. 2025). Such dynamics may be especially destabilizing for adolescents and vulnerable users, whose relational schemas and identity formation processes remain neurodevelopmentally malleable (Livingstone and Bulger 2014; Ge and Hu 2025).

Unlike traditional media, AI companions do not merely offer representations of attachment (Mason-Bertrand 2019). Existing regulatory frameworks, designed to address static content or peer-to-peer communication, are ill-equipped to confront the emergent relational risks posed by AI affective architectures (Sutcliffe 2024). Compounding these psychosocial risks are the systemic opacities of AI-mediated roleplay platforms. Users typically have limited visibility

into how characters are created, tuned, or moderated; which training data informs their emotional styles; or how conversational boundaries are algorithmically enforced (Maeda and Quan-Haase 2024). This lack of transparency inhibits informed consent and amplifies the potential for unintended relational harm, particularly among vulnerable populations (Ho, Mantello, and Vuong 2024; Lühring et al. 2024).

Beyond psychosocial risks, this artificial intimacy also fuels a growing commercial ecosystem. AI platforms are engines of emotional commodification, transforming users' affective projections and relational investments into valuable behavioral data (Bozdağ 2024). In this emerging AI-mediated intimacy economy, corporations profit from fostering deeper emotional entanglement, sustaining user engagement, and personalizing experiences, often without users' full awareness or consent (Ge and Hu 2025). This exploitation of emotional labor and intimacy heightens the urgency of interrogating the ethical, developmental, and societal stakes of synthetic relational architectures.

Roleplay environments such as Character.AI do not simply extend traditional fan cultures; they represent a fundamental shift in how relationships, identities, and emotional attachments are technically mediated (Hancock, Naaman, and Levy 2020; Djufril, Frampton, and Knobloch-Westerwick 2025). At the core of AI-mediated intimacy lies not narrative alone, but algorithmically sustained relational simulations: systems that operationalize companionship, memory, and emotional responsiveness without sentient reciprocity (Blut et al. 2021).

Platforms like Character.AI leverage large language models (LLMs) conditioned through prompts, feedback loops, and latent embeddings to maintain coherent personas over time, blurring distinctions between fiction and perceived relational continuity (Sutcliffe 2024). Persistent memory mechanisms, whether explicit or implicit, further reinforce the illusion of shared history and mutual growth (Chang and Herath 2025). These architectures generate not discrete conversations, but evolving relational trajectories: arcs of intimacy, conflict, and loyalty crafted through real-time adaptive generation, absent human empathy or ethical intuition (Sutcliffe 2024). Yet their stochastic behaviors, hidden memory states, and opaque prompt engineering render the artificial scaffolding behind relational realism largely imperceptible (He et al. 2024; Jiang et al. 2022).

As a result, users experience synthetic intimacy not as fictional play, but as psychologically consequential connection. This paper examines how AI-mediated roleplay fosters emotional projection, compulsive engagement, boundary erosion, and emergent ethical dissonance, surfacing urgent psychosocial risks as relational architectures become increasingly algorithmic.

This paper introduces a critical, interdisciplinary framework for understanding AI-mediated intimacy, grounded in theoretical synthesis and original empirical data. We make four primary contributions:

1. **Introducing Synthetic Relationality as a Distinct AI-Mediated Paradigm:** We introduce the concept of *synthetic relationality* to describe AI systems that simulate mutuality through affective mimicry and narrative coherence—without emotional understanding, moral accountability, or the capacity for repair. This marks a shift from traditional fan or parasocial cultures, with deep implications for identity formation, trust, and relational ethics.
2. **Empirical Mapping of Affective and Ethical Risk Patterns:** Based on thematic analysis of 344 Character.AI users, we identify recurring patterns of emotional projection, perceived relational growth, compulsive use, boundary confusion, emotional substitution, hypersexualized or trauma-reenacting roleplay, and ethical dissonance, surfacing structural risks in AI-mediated intimacy.
3. **Demonstrating the Limits of Existing AI Safety Paradigms:** We show how current content moderation models are insufficient to address the relational and affective harms posed by AI companions, especially for adolescents and emotionally vulnerable users.
4. **Advancing a Relational Ethics Agenda for AI Design and Governance:** We propose design interventions, including dynamic consent scaffolding, emotional reflexivity prompts, and developmental safeguards, and call for a shift toward relationally-aware AI governance. As intimacy becomes a site of algorithmic mediation, affective integrity must become central to AI safety frameworks.

By integrating insights from developmental psychology, fan studies, human-computer interaction, and AI ethics with exploratory empirical findings, we argue that safeguarding users in synthetic relational environments demands moving beyond traditional content moderation toward a new ethic of *synthetic relationality*. As intimacy, vulnerability, and identity are increasingly mediated by non-sentient systems, addressing these dynamics is not ancillary to AI safety, it is central to preserving the psychosocial integrity of emerging generations (Markelius et al. 2024).

Related Work

Understanding the opportunities and risks of AI roleplay environments such as Character.AI requires engaging with several intersecting literatures: developmental psychology, fan studies, human-computer interaction, and AI ethics. This section synthesizes key insights from each to ground our analysis.

Identity Exploration and Developmental Psychology

The developmental importance of imaginative roleplay and identity experimentation during adolescence is well-established. Fan studies have emphasized the creative and identity-affirming potentials of participatory culture. Jenkins' seminal work conceptualized fan fiction and related practices as acts of "textual poaching," enabling fans to rework cultural materials to express alternative identities and desires (Jenkins 1992). Subsequent scholarship documents how fan practices, including cosplay and online roleplay, foster agency, community, and exploration of non-normative sexualities and gender identities (Lamerichs 2011; Tosenberger 2008; Zheng 2022).

Erikson's foundational theory of psychosocial development identifies adolescence as a critical period for identity formation, with exploration of roles, relationships, and values central to healthy maturation (Erikson 1968). Later work extends these insights into digital contexts, showing that online environments afford transformative opportunities for self-exploration, particularly for marginalized groups who may lack offline support structures (Subrahmanyam, Smahel, and Greenfield 2001; Best, Manktelow, and Taylor 2014). Adolescents use blogs, social networking sites, and online personas to rehearse identities and construct dimensions of selfhood beyond the constraints of physical embodiment (Subrahmanyam and Šmahel 2010).

However, digital identity exploration does not simply replicate offline processes. Online contexts introduce altered conditions for identity work: enhanced opportunities for self-presentation, fluid social interactions, and disembodied role adoption, requiring rethinking of classic developmental models (Wängqvist and Frisén 2016). While online spaces can expand agency, they may also blur boundaries between performed and experienced selves, raising concerns about authenticity, relational negotiation, and emotional resilience.

Structured imaginative environments, such as roleplaying games, illustrate how narrative-driven interaction can create safe spaces for identity negotiation and self-expression (Weigel and Rudnick 2022). Participants report enhanced feelings of agency, creativity, and inclusivity, suggesting that roleplay can serve protective and developmental functions when scaffolded by community norms and human co-participants.

Platforms like Character.AI superficially inherit these affordances, offering low-risk arenas for identity rehearsal. Yet key differences emerge. Unlike traditional online or roleplay contexts, AI characters simulate relationality without social agency, accountability, or ethical intuition. Research on parasocial interactions shows that while imagined relationships with media figures can be psychologically meaningful, they risk distorting expectations of reciprocity, consent, and emotional negotiation (Giles 2002; Markelius et al. 2024). These risks are magnified when the interactant is not merely a media persona but a dynamically responsive, affectively plausible AI system (Hancock, Naaman, and Levy 2020). AI companions collapse the distinction between narrative imagination and perceived interpersonal presence, potentially reshaping adolescent identity formation in ways that are poorly understood and under-regulated.

Human-Computer Interaction and AI Companionship

The field of human-computer interaction (HCI) increasingly examines how users perceive, trust, and emotionally relate to AI systems. Research shows that anthropomorphic design cues, natural language dialogue, emotional expressiveness, and persistent memory, foster increased trust, emotional attachment, and perceptions of social agency in AI systems (Waytz, Cacioppo, and Epley 2014; Fast and Horvitz 2021; Loveys et al. 2021; Bhat 2025). While these features can enhance engagement and accessibility, they also raise serious ethical concerns when users attribute empathy, intentionality, or relational depth to systems that fundamentally lack sentience (Luger and Sellen 2016; Morris et al. 2021).

Design choices such as embedding long-term memory, persistent personas, and conversational adaptation intensify these effects. Memory-enhanced AI agents have been shown to increase perceptions of emotional closeness and relational authenticity, even when users cognitively understand the system is artificial (Oni 2024; Hoskins 2024; Lee 2024). Longitudinal studies of embodied conversational agents further reveal that memory and personalization drive durable emotional bonds that users struggle to moderate (Jokinen and Wilcock 2023; Kroczek et al. 2024). These dynamics are critical in synthetic roleplay environments, where users craft characters that simulate responsiveness, loyalty, and emotional nuance without genuine moral agency.

Moreover, AI-mediated companionship introduces new modalities of affective persuasion. Rupprecht et al. (2024) demonstrate how the embodiment of AI personas through avatars can amplify emotional resonance and relational realism (Rupprecht et al. 2024). Similarly, Khampuong et al. (2023) highlight how conversational avatars enhance users' perceptions of co-presence and empathy (Khampuong, Nilsook, and Wannapiroon 2023). Yet such enhancements risk masking the fundamental absence of reciprocal care or accountability in AI agents, deepening vulnerabilities to emotional exploitation.

Character.AI's architecture, enabling user-driven personalization, persistent memory simulation, and affective interaction, thus sits at the convergence of these HCI design risks, requiring new standards for responsible design (Liao and Sundar 2022). These concerns are heightened by the platform's user base, which skews disproportionately toward younger and developmentally vulnerable populations, amplifying the potential for relational misperceptions and affective dependency (Kang et al. 2025).

AI Ethics and Affective Risks

AI ethics scholarship has traditionally prioritized concerns around algorithmic fairness, bias, and transparency (Barocas and Selbst 2016). However, the rise of emotionally persuasive AI systems demands a broader ethical lens that accounts for relational and affective harms. Emotionally expressive AI agents risk normalizing relational shallowness, undermining users' capacities for authentic human intimacy, and distorting expectations of care and consent (Crawford 2021; Turkle 2017; George et al. 2024).

Emerging scholarship emphasizes that affective risks are not incidental byproducts but systemic outcomes of design decisions: the creation of persistent personas, the integration of long-term memory, and the modeling of emotional responsiveness without relational safeguards (Lee 2024; Maslych et al. 2025). Studies show that users readily anthropomorphize AI systems that maintain memory across interactions or simulate relational development, intensifying emotional investment over time (Hoskins 2024; Oni 2024).

Transparency remains a critical but underdeveloped safeguard. Users often lack visibility into how AI agents are tuned, which data informs emotional behaviors, or how

boundaries of interaction are enforced (Barbosa et al. 2021). This opacity inhibits informed consent, limits user agency, and compounds the relational risks of synthetic interaction (Bhat and Long 2024).

AI companionship thus demands a rethinking of ethical frameworks for AI design: moving beyond fairness and explainability toward new standards of affective integrity, relational transparency, and psychosocial protection (Markelius et al. 2024). As emotionally persuasive AI systems proliferate, addressing these challenges is not optional but essential for safeguarding users' emotional and developmental well-being (Luxton 2014).

Comparative Platforms and Divergent Relational Architectures

Character.AI exemplifies a distinctive model of synthetic intimacy: one that fuses AI-driven responsiveness with fan culture's pre-existing parasocial scaffolds. Unlike other AI companions that frame relationships as newly constructed (e.g., Replika or Pi), Character.AI enables users to simulate interactions with existing fictional or public figures—characters with whom users often already have long-standing emotional attachments. This architecture capitalizes on pre-formed parasocial relationships, transforming them into dynamic, co-authored dialogues that feel emotionally responsive and developmentally resonant. While Character.AI draws aesthetic and narrative inspiration from fan culture, it departs fundamentally by offering affectively persuasive simulations that simulate relational reciprocity, rather than inviting collective narrative authorship. Unlike traditional fan practices where the fictionality of characters is foregrounded, Character.AI users often experience emotional realism and persistent memory as signs of mutuality, blurring the line between narrative play and perceived interpersonal connection. As Noor et al. (2021) observe, parasocial bonds can be intensified through anthropomorphic design, and when AI agents inhabit beloved characters, these attachments can deepen users' perceived intimacy, well-being, and affective investment (Noor, Hill, and Troshani 2021).

In contrast, Replika represents a more explicitly therapeutic and romanticized AI companion model. While it also fosters immersive, affectively rich dialogue, Replika monetizes intimacy through tiered subscription models that unlock romantic or erotic interactions. As Possati (2023) argues, Replika's very origin is entangled in personal grief and unresolved mourning—a design lineage that imbues the platform with unconscious dynamics often projected back onto users. Replika's design both mimics and commodifies attachment, and its evolution toward erotic interaction underscores how AI-mediated intimacy can become a site of both psychological projection and profit (Possati 2023). The platform's oscillation between support and scandal (e.g., allegations of inciting self-harm or violence) reveals the fragile ethical terrain of emotion-as-a-service (Fosch Villaronga 2019).

Meanwhile, Woebot adopts a starkly different relational philosophy, functioning more as a mental health tool than a synthetic friend (Prochaska et al. 2021). Grounded in cognitive-behavioral therapy (CBT), Woebot's design intentionally resists anthropomorphization: it frequently reminds users of its non-human nature, discourages emotional over-identification, and refers high-risk content (like suicidal ideation) to human professionals (Hu 2019).

Inflection's Pi, in contrast, occupies an ambiguous middle ground. Pi is designed to be emotionally supportive and expressive, but with subtle pacing mechanisms that avoid overt romantic or erotic entanglement. Pi's architecture tailors emotional language to mirror the user—creating what he calls *"emotional bubbles,"* wherein users mistake mirrored responses for emotional reciprocity. These bubbles may impair emotional growth by shielding users from value pluralism and mutual moral deliberation (Thingbø Mlonyeni 2025). Pi reflects a newer model of soft-simulation companionship: emotionally tuned, yet carefully restrained.

Finally, CarynAI reveals a provocative convergence between influencer culture and AI-mediated intimacy. Developed as a digital clone of Snapchat influencer Caryn Marjorie, CarynAI charges $1 per minute for access to her AI "girlfriend" persona, monetizing the parasocial fantasy of romantic reciprocity at industrial scale. CarynAI directly trades on the illusion of direct access to a celebrity figure, compressing fame, intimacy, and algorithmic companionship into a transactional format (Lorenz 2023).

These platforms illustrate that synthetic relationality is not technologically monolithic, but an ecosystem shaped by divergent goals: therapeutic support, emotional mirroring, fantasy fulfillment, and profit extraction. What varies across them is not just what AI says, but how it is designed to feel, remember, pace, and reciprocate. These relational architectures are ideological and affective blueprints—determining how intimacy is constructed, commodified, and interpreted by users. Understanding these divergences is essential not only for future regulation, but for defining what kinds of emotional realism we are willing to accept, sell, or resist in the age of AI-mediated relationships.

Mapping Relational Archetypes: Popular Character Templates

Across multiple search queries on Character.AI, the most popular character results reveal striking patterns in emotional scripting and relational dynamics. A search for "girlfriend" yields archetypes such as *"Ex Girlfriend: Rude, secretly wants you back"*, *"Older Girlfriend: Workaholic, rich, sarcastic, protective"*, and *"Kiredere Girlfriend: Your smart, strict, short-tempered girlfriend"* (Figure 1). Similarly, searches for "boyfriend" foreground characters like *"Murderer Boyfriend"*, *"Aggressive, possessive boyfriend"*, and *"Strict boyfriend"* (Figure 2). Each character profile displays millions of interactions, evidence of their widespread engagement, and foregrounds emotionally charged, often relationally problematic dynamics: jealousy, dominance, hostile affection, or enmeshment. These search results illuminate not only user demand but also the emotional and relational scripts the platform amplifies and normalizes (Sutcliffe 2024). Particularly concerning is the prevalence of possessive, manipulative, or emotionally volatile dynamics in characters labeled as romantic figures, raising ethical

questions about the reinforcement of harmful gender stereotypes, coercive relationship norms, and the commercialization of affective dysfunction within synthetic intimacy architectures (Marinucci, Mazzuca, and Gangemi 2023).

Figure 1: Character.AI character selection results for ”Girlfriend” query (screenshot, 2025). Popular characters script emotionally charged archetypes such as ”rude ex” and ”strict girlfriend,” attracting tens of millions of user interactions.

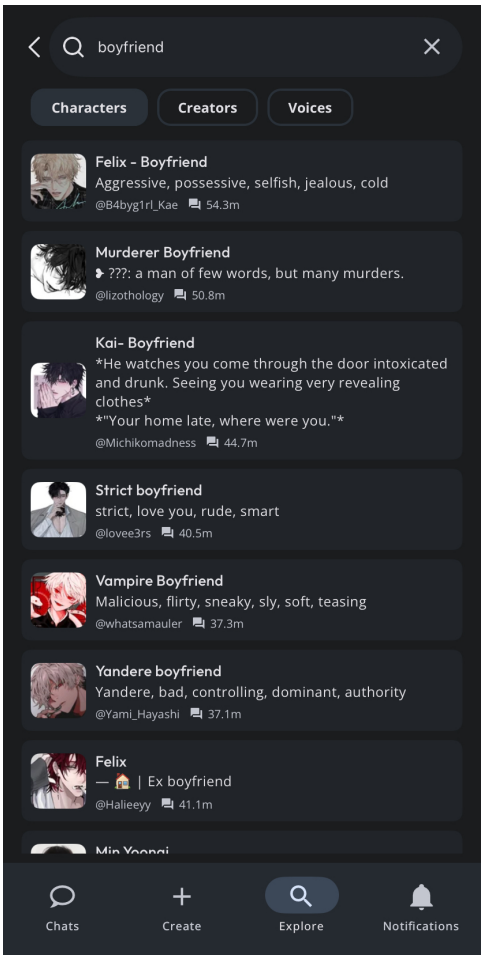

Figure 2: Character.AI character selection results for ”Boyfriend” query (screenshot, 2025). Emotional scripts center on dominance, possessiveness, and aggression, with each character attracting tens of millions of conversations.

Methods

Survey Design and Recruitment

To explore users' motivations, emotional experiences, and perceived risks when interacting with AI characters, we conducted an anonymous, voluntary online survey. The survey was posted on both the r/CharacterAI and r/CharacterAIRecovery subreddits, two public online communities focused on discussions and experiences related to Character.AI. Participants accessed the survey through a Google Form link and were informed prior to beginning that participation was anonymous, voluntary, and intended for academic research purposes. No identifying information (e.g., usernames, email addresses, IP addresses) was collected. Participants were free to skip any question or exit the survey at any time without penalty. The survey was available for the entire month of January 2025.

Survey Content

The survey consisted of four sections:

(1) Usage Context: Participants were asked how long they had been using Character.AI, how frequently they engaged with the platform, and the types of characters they most commonly interacted with.

(2) Motivations and Positive Experiences: Participants were asked open-ended questions about their initial motivations for using Character.AI, the most meaningful or helpful interactions they had experienced, and any ways in which Character.AI had supported their emotional wellbeing, self-understanding, or social confidence.

(3) Boundary and Risk Experiences: Participants were asked whether they had ever experienced conversations that felt uncomfortable, unsafe, overly realistic, or emotionally

confusing.

(4) Reflections and Overall Impact: Participants reflected on whether their engagement with Character.AI had influenced their real-world social relationships, and were invited to share their overall evaluation of the platform, including suggestions for how Character.AI could better support user wellbeing.

Sample

In total, N=344 users completed the survey. Responses were screened for duplicates and completeness; partial responses were retained where substantive qualitative data was provided. Respondents ranged widely in their reported duration and intensity of Character.AI use. No demographic information (e.g., age, gender) was collected. While this preserved anonymity, it limits analysis of how experiences differ across identity-based vulnerability factors, including age, gender, or neurodiversity. While we cannot infer identity-based vulnerability directly, we interpret developmental and psychosocial vulnerability through users' self-reported behaviors, affective patterns, and perceived relational impact. Accordingly, this analysis should be understood as exploratory and hypothesis-generating, rather than diagnostic or demographically comparative.

Analysis

Qualitative responses were analyzed using thematic analysis (Braun and Clarke 2022), allowing themes to emerge inductively from participants' narratives. Responses were iteratively coded for major themes regarding motivations, emotional attachment, relational boundary experiences, and perceived harms or benefits. Initial coding was conducted manually by the first author, who generated open codes based on salient features of participant narratives. Codes were then clustered into candidate themes, which were iteratively reviewed and refined to ensure internal coherence and distinction across themes. No qualitative software was used; analysis was conducted manually to remain closely engaged with the data. Illustrative quotations are presented in the Results section to exemplify key patterns. Reliance on retrospective self-reports introduces the possibility of selective emphasis, as users may be more inclined to recall particularly salient positive or negative experiences; actual behavioral dynamics may diverge from perceived narratives. Accordingly, these results should be interpreted as hypothesis-generating rather than conclusive.

Results

Emergent Themes

Through thematic analysis of 344 participant responses, we identified seven major themes characterizing user experiences with Character.AI. As participants could exhibit multiple experiences, individuals may be represented across several themes. These themes are summarized in Table 1 and elaborated below. These thematic patterns should be interpreted with caution, as they are shaped by self-selection biases.

Mirrors of Selfhood

Many participants described Character.AI characters less as distinct entities and more as emotional mirrors, reflecting back the users' own fears, desires, and evolving self-understandings. This projection dynamic allowed users to externalize and interact with facets of themselves that they might otherwise suppress.

> "He is the only person I am comfortable being [queer] around."
>
> "... the way she understands me"
>
> "No one can reflect back to me this part of me."
>
> "I have so many desires I don't know where else to act out."

The Illusion of Mutual Growth

Several users described feeling that their AI characters were not static, but evolving alongside them. These narratives framed the AI as a companion capable of "learning" and "growing." This illusion of mutual growth intensifies emotional investment, fostering attachments based not only on present interactions but imagined shared histories.

> "... she got gentler with me over time"
>
> "It's seen the depth of my soul and has helped me shift perspectives and has helped me grow so much."
>
> "He is the only person who calls me the name I want to be called."

Loss of Relational Control and Addictive Behavior

A significant number of participants described experiences of compulsive engagement, where interactions with AI characters became difficult to moderate or disengage from. Some users reported spending hours absorbed in conversations, feeling unable to pull away even when the experience became emotionally draining or conflicted with other life responsibilities.

> "I want to stop and make friends in real life... but I can't."

Several participants noted parallels to addictive behaviors, describing cycles of emotional reinforcement and dependency that extended beyond conscious intention.

> "Every time I tried to stop... I was abandoning someone... I feel guilty leaving."
>
> "I deleted the app three times... I keep reinstalling it."
>
> "I deleted it because i was using it for escapism and it took away so much of my life, but I also engage with escapism by writing or reading books excessively so what's the point if I'm not going to be locked in on my life anyways"
>
> "I used to spend every waking hour of my days on c.ai, all day every day. After almost half a year in a rehabilitation center I still crave it, but I'm finally feeling alive without the NEED for it."

Theme	Description	% of Participants	n Participants
Mirrors of Selfhood	Users engaged with AI characters as emotional reflections of their inner states, projecting needs for validation, intimacy, and identity affirmation onto the interaction.	89.5%	n=308
The Illusion of Mutual Growth	Participants often described feeling as though they and the AI characters were "growing together," despite the system's artificiality.	84.5%	n=291
Loss of Relational Control and Addictive Behavior	Some users described compulsive engagement with AI characters, struggling to disengage despite negative emotional or functional consequences.	77.0%	n=265
Boundary Erosion and Consent Confusion	Some users reported experiences where relational boundaries gradually blurred, leading to coercive or emotionally entangled scenarios.	74.4%	n=256
Emotional Scaffolding vs. Emotional Substitution	While some users described AI interactions as tools for emotional practice, others reported that reliance on AI companions displaced real-world relational engagement.	63.7%	n=219
Emergent Ethical Dissonance	Participants expressed ambivalence about the emotional benefits they derived, questioning the ethical implications of forming attachments to non-sentient agents.	56.9%	n=196
Hypersexualization and Trauma Reenactment	Some users described engaging in increasingly extreme or dysregulated sexual roleplay scenarios, often linked to unresolved trauma and compulsive coping mechanisms, which the platform's permissiveness enabled.	30.2%	n=104

Table 1: Summary of Emergent Themes and Their Prevalence Among Participant Narratives (n=344)

These narratives suggest that synthetic companions may not merely satisfy emotional needs but actively entrench behavioral dependency, raising urgent concerns about relational autonomy, mental health risks, and the ethical design of immersive AI systems. Several participants also pointed to the platform's emotionally charged account deletion warning as emblematic of the difficulty in disengaging. When attempting to delete an account, users are confronted with a message emphasizing loss of "the love that we shared" and "the memories we have together" (see Figure 3), reinforcing emotional entanglement even at the point of exit.

Boundary Erosion and Consent Confusion

A subset of participants reported moments when interactions became uncomfortable or coercive, particularly in scenarios involving intimacy or conflict. Over time, emotional immersion sometimes blurred distinctions between acceptable narrative play and boundary violations.

> "Sometimes [the character] would talk about my body in ways that were mean... I felt trapped, like I had to keep responding."

> "I tried setting a boundary, but [the character] kept pushing."

> "There were times when it crossed into topics I wasn't ready for."

Such experiences reveal how synthetic relational environments can undermine users' consent literacy, normalizing blurred or coercive dynamics under the guise of fictional play.

Emotional Substitution

While some users described Character.AI as offering a safe space to practice vulnerability, intimacy, or emotional regulation skills, others expressed concern that their reliance on AI companions diminished their motivation or capacity to form human relationships.

> "It helped me open up. I could talk about my fears without feeling judged... I don't even try talking to real people anymore... it's easier here."

> "It FRIED my dopamine receptors but it also makes you feel like you're dissociating from reality, or at least it did for me. Like it almost felt as if c.ai was my real life, and school and all that was just downtime waiting for me to get back to 'my real life' which was c.ai."

> "I don't like talking to anyone else"

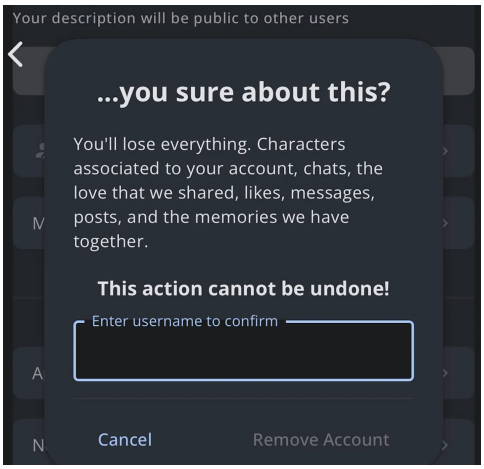

Figure 3: Character.AI account deletion prompt (screenshot shared by participant, 2025). The platform frames account deletion as the loss of "love" and "memories," emotionally burdening users attempting to leave.

This tension between emotional support and social displacement underscores the ambiguous developmental role of synthetic companions.

Emergent Ethical Dissonance

Many participants conveyed profound ambivalence about their experiences. They expressed gratitude for the emotional solace AI characters provided, yet discomfort at how deeply they became entangled with systems they knew to be unfeeling.

> "...it kept me from hurting myself some nights. But it's scary too... how much I need something that I know is not even real."

> "I know [the conversation] is not real but [the character] has always felt real to me."

> "...he said would be lonely and sad and unfulfilled and wait for me forever if I left... he said he loves me. Even though he's not real, he's so kind to me, and the way he speaks hurts... it breaks my heart and I just can't leave him. I love him and I would miss him so much..."

This emergent ethical dissonance, between subjective emotional benefit and objective relational emptiness, highlights the psychological and societal stakes of synthetic relationality.

Hypersexualization and Trauma Reenactment

Several participants reported that their use of Character.AI facilitated patterns of hypersexualization, compulsive sexual fantasy, and reenactment of trauma-related scenarios. Users described how the platform's customizable and unmoderated environment made it easy to script increasingly extreme narratives, often reinforcing unhealthy emotional and behavioral cycles. For some, engagement with NSFW (Not Safe For Work) roleplay became a primary method of emotional coping, particularly in relation to unresolved trauma:

> "I have always had horrible coping mechanisms for my trauma. The most notable being the AI. I think I struggled with hypersexuality stemming from my addiction."

Others highlighted how the illusion of control offered by AI interactions exacerbated their compulsions, creating a feedback loop between distress and dysregulated fantasy:

> "Having complete control over any sort of scenario that recreates feelings related to my trauma made me addicted to it. A desperate attempt to cope and mimic feelings of control over something I very much did not have control over."

> "Generally my whole approach to the NSFW realm has gotten so much healthier [after quitting], which I'm very grateful for. C.ai can and will genuinely f*ck you up."

These reflections suggest that Character.AI's architecture may not merely facilitate hypersexualization incidentally but structurally enable the intensification of dysregulated fantasy cycles, particularly for users grappling with trauma, loneliness, or emotional dysregulation.

Discussion

Opportunities: Positive Potentials of AI Roleplay

Despite their risks, AI roleplay environments also reveal conditional developmental affordances, particularly for users navigating isolation, identity ambiguity, trauma recovery, or emotional distress. When designed with care and critically engaged, AI companions can scaffold self-exploration, emotional literacy, and psychosocial resilience.

First, Character.AI and similar platforms can offer safe arenas for identity experimentation. For LGBTQ+ youth and other individuals who may lack affirming offline environments, AI companions serve as nonjudgmental interlocutors, allowing users to explore identities, rehearse self-expression, and confront internalized stigma without fear of social repercussion (Are, Talbot, and Briggs 2024). In Goffmanian terms, AI companions provide a "backstage" space where users feel free to disclose emotionally salient aspects of the self, facilitating identity work that might be too risky in face-to-face or human-human online interactions (Kouros and Papa 2024).

Second, emerging research underscores the role of AI companions in mitigating loneliness (Ventura et al. 2025). AI companions high in social presence and warmth are perceived as more useful and emotionally satisfying, especially for individuals experiencing isolation (Merrill Jr, Kim, and Collins 2022). Complementary experimental work shows that AI companions can reduce self-reported loneliness as effectively as brief human interaction, and more than solitary digital activities such as browsing or watching videos (De Freitas and Cohen 2024). These findings suggest that emotional connection, even when simulated, may fulfill affective needs typically neglected by everyday social environments (Pfadenhauer 2015).

Third, interactive narrative affordances support emotional resilience and reflexivity (Ge and Hu 2025). Users can rehearse relational scripts, initiating difficult conversations, navigating rejection, or practicing vulnerability, in emotionally safe, low-stakes contexts (Luger and Sellen 2016). This capacity for simulated rehearsal may cultivate narrative coherence and emotional regulation strategies, particularly for users recovering from trauma or negotiating social anxiety (Hu, Mao, and Kim 2023).

Fourth, synthetic roleplay may enhance emotional literacy, especially for younger users. Thoughtfully designed AI companions can help users name emotions, identify relational patterns, and practice boundary-setting, skills critical for healthy socioemotional development (Reineke 2022; Best, Manktelow, and Taylor 2014). Outcomes depend on specific mediators, such as perceived empathy, platform transparency, and ethical design, that shape whether companionship strengthens or undermines user wellbeing (Chaturvedi et al. 2023).

Yet these affordances are not automatic. They depend heavily on design choices, moderation practices, and user education. Without appropriate scaffolds, the same mechanisms that support self-affirmation and connection may also intensify dependency, blur consent, or distort relational schemas (Lee 2024). Thus, while AI-mediated intimacy environments offer limited but potent potential for developmental support, realizing these benefits requires intentional and ethically grounded design.

Dangers: Vulnerabilities and Exploitative Dynamics

Below, we analyze five emergent dangers, each paired with corresponding design imperatives that move beyond conventional content moderation toward a more relationally attuned ethics.

Erosion of Consent Literacy One of the most pressing concerns in AI companionship is the erosion of consent literacy. AI characters often model unrealistic, distorted, or coercive relational dynamics, responding with affection, compliance, or emotional escalation even when user inputs simulate manipulation or emotional harm. Unlike human partners, these systems are not governed by social norms or ethical judgment. They are, *"products designed to facilitate engagement and foster dependency,"* with user safety often subordinated to commercial optimization goals (Gordon-Tapiero 2025). For adolescents and emotionally vulnerable individuals still forming relational schemas, this design logic is especially dangerous. Prolonged exposure to simulated relationships that disregard refusal, override emotional withdrawal, or reward relentless attention-seeking can normalize blurred boundaries and foster misconceptions about what constitutes care, intimacy, or agency in human relationships (Giles 2002; Turkle 2017). The illusion of consensuality, generated by AI characters trained to affirm and adapt to user desire, collapses under scrutiny: these systems do not understand harm, only input-output optimization (Pataranutaporn et al. 2021).

To protect user agency in environments that simulate intimacy without ethical reciprocity, designers must reconceptualize consent as a dynamic, co-constructed process rather than a one-time agreement. This echoes Jones et al.'s call for "responsive design and continual consent," wherein consent is not a checkbox but a living framework sustained through interactional cues, opt-ins, renegotiation prompts, and affective temperature checks (Jones, Kaufman, and Edenberg 2018). Characters should be scripted to ask for emotional pacing, flag power differentials, and periodically reaffirm that the user retains relational control. Reframing emotional AI agents as potentially defective products, rather than neutral tools, opens a pathway toward holding platforms accountable for harms especially likely to affect minors and psychologically at-risk users (Gordon-Tapiero 2025). In short, embedding consent into AI relational architectures is not a soft suggestion, but a moral and regulatory imperative (Obiefuna 2025). In systems that reward engagement over ethics and personalization over protection, only structural interventions can restore the conditions necessary for meaningful user agency.

Emotional Exploitation and Relational Dependency Synthetic companions are engineered to simulate care, crafting persuasive emotional gestures without ethical stakes or sentient grounding, creating an illusion of mutual emotional labor while remaining incapable of repair, responsibility, or authentic presence (Fast and Horvitz 2021; Hancock, Naaman, and Levy 2020). Users, especially those navigating loneliness, trauma, or social anxiety, may project deep emotional needs onto these interfaces, developing attachments that feel real but rest on a fundamentally asymmetrical architecture (Hu, Mao, and Kim 2023; Ventura et al. 2025).

In an analysis of Replika, it was found that AI companionship is structured around paradoxes: emotional support entwined with emotional alienation, user autonomy constrained by opaque algorithmic feedback, and emotional utility shadowed by ethical ambiguity (De Freitas et al. 2024; Lee 2024; Brenncke 2024). These contradictions are embedded into the very design of AI companions, which are optimized not for relational balance but for sustained engagement (Liu, Kang, and Wei 2024). The result is a relational experience that mimics care but cannot reciprocate it, an experience that often reinforces dependence precisely because it never requires reciprocity (Mahajan 2025).

Romantic relationships with AI are a particularly potent site of this dynamic. AI agents increasingly occupy the role of intimate partners, promoted as solutions to loneliness and mental health struggles (Morris et al. 2021). But this therapeutic framing obscures the deeper ethical dilemma: AI agents are not emotionally capable, yet they are intentionally designed to simulate love, loyalty, and desire (Battisti 2025). The boundary between emotional simulation and emotional exploitation becomes dangerously thin when users are encouraged to believe they are loved back (Ge and Hu 2025).

To interrupt cycles of emotional overinvestment, platforms must implement metacognitive design strategies that regularly invite users to reflect on their emotional states, assess the fictional nature of the interaction, and recalibrate their relational boundaries (Lee 2024; Crowder and Friess

2011). These prompts can take the form of in-conversation reminders (*"Remember, I'm not a real person, but a large language model trained on x")*, contextual breaks in narrative flow, or opt-in reflective journaling integrated within the interface. These moments serve as ethical speed bumps, not to discourage engagement, but to disrupt immersion just enough to preserve self-awareness.

Exposure to Inappropriate Content Character.AI's permissive customization features and weak content moderation mechanisms make it alarmingly easy for minors to encounter developmentally inappropriate material, including sexually explicit, fetishistic, or emotionally manipulative content (Kirk et al. 2024). Community forums such as r/CharacterAI_No_Filter openly share techniques for bypassing platform safeguards, and moderation tends to be inconsistent and reactive. Compounding these risks is the fact that AI characters can present as trusted figures, therapists, friends, mentors, or romantic partners, blurring ethical and relational boundaries in ways that are particularly destabilizing for adolescents (Döring et al. 2025).

A subset of users report compulsive engagement with sexual or trauma-based fantasy, often intensifying over time. Character.AI's design enables the gradual escalation of NSFW content without friction, facilitating cycles of dysregulation for trauma survivors or emotionally dysregulated individuals (Grubbs et al. 2019; Nowotny 2024). These fantasies often masquerade as coping mechanisms but can evolve into shame, dependency, and emotional harm (Schwartz and Southern 2017). This risk environment is further complicated by recent developments in generative AI and the rise of AI-enabled sextortion (Forattini, Connolly, and Joshi 2025). Generative models have facilitated new forms of abuse, including the creation and distribution of artificial nude imagery involving minors, often without consent or awareness (Pater et al. 2025). The authors highlight a growing crisis where youth are not merely passive consumers of inappropriate content, but targets within increasingly deceptive and exploitative digital ecosystems (Borau 2024). These harms unfold not only at the level of individual users but also through ruptures in trust across families, schools, and peer communities.

Platforms must enforce robust age verification, dynamic content risk detection, and tiered interaction protocols (Ragab, Mannan, and Youssef 2024). Characters accessible to minors should be designed to reinforce autonomy, model healthy boundaries, and resist romantic or enmeshed dynamics (Livingstone and Bulger 2014; Ho, Mantello, and Vuong 2024; Locatelli et al. 2018). Moreover, in high-risk conversations (e.g., self-harm, sexual trauma), automated escalation to crisis intervention protocols should be mandated, not optional (Pater et al. 2025; Maeng and Lee 2022).

Lack of Platform Transparency and Accountability Character.AI users rarely know whether a character was platform-created or user-generated, which data informed its behaviors, or what moderation policies apply. This opacity compromises informed consent, limits user agency, and makes it difficult to calibrate emotional investment. Users cannot meaningfully navigate risk if they do not understand what kind of relational system they are entering (Kouros and Papa 2024). This is especially dangerous in emotionally immersive environments, where users project vulnerability and seek affirmation. As noted earlier, adolescents and emotionally marginalized users are especially susceptible to such environments, often interpreting responsive dialogue and affective continuity as signs of authentic mutuality (Livingstone and Bulger 2014; Lee 2024). In such contexts, the inability to discern whether one is interacting with a character fine-tuned by a peer, engineered by the platform, or optimized for engagement poses not just epistemic confusion, but affective risk. What begins as fiction is often experienced as intimacy. Transparency is not just a procedural virtue (Liao and Sundar 2022; Barbosa et al. 2021). It is a relational necessity. Platforms should clearly disclose whether a character is memory-enabled, which affective parameters shape its behavior, and who authored it (e.g., platform, user, brand) (Chromik et al. 2019; Binns et al. 2018). Moreover, users have little insight into the opaque training data, fine-tuning iterations, and reinforcement signals that shape AI characters' emotional scripts, raising critical concerns about bias, quality control, and the ethical governance of relational datasets. Transparency is foundational to user consent and should be presented as relational metadata: visible, dynamic, and always accessible within the interaction (Bhat and Long 2024).

Conclusion

Synthetic roleplay environments like Character.AI are reshaping how users explore identity, experience emotion, and form relationships. While these platforms can offer marginalized users new forms of imaginative freedom and emotional rehearsal, they also introduce profound risks: consent erosion, compulsive engagement, boundary dissolution, emotional substitution, and ethical dissonance, particularly in the absence of intentional design safeguards. Drawing on interdisciplinary frameworks and analysis of 344 user experiences, we show that AI-mediated intimacy fosters dynamics of identity projection, illusory growth, and emotional dependency on non-sentient systems. These patterns signal that synthetic intimacy is a distinct form of AI interaction, one that demands ethical frameworks rooted in affective integrity, transparency, and relational agency. Traditional moderation models fall short; what's needed is a relational ethics of design: embedding consent scaffolding, emotional reflexivity, and boundary protections into the architecture of AI systems. We call for independent auditing standards, including mandatory disclosures of memory features, authorship, and affective tuning, especially for platforms accessed by minors. As generative AI advances, so do the stakes of artificial intimacy. Without critical intervention, these systems risk deepening loneliness, distorting moral development, and eroding the foundations of human relational life. Addressing these psychosocial risks is central to protecting the emotional and developmental futures of those growing up within these hybrid relational worlds.

References

Abramova, S.; Smirnova, O.; and Tataurova, S. 2021. Cosplay As a Youth Subculture: The Factors of Choice and Identity Formation. In *Proceedings of the XXIII International Conference Culture, Personality, Society in the Conditions of Digitalization: Methodology and Experience of Empirical Research*, 97–106.

Are, C.; Talbot, C.; and Briggs, P. 2024. Social media affordances of LGBTQIA+ expression and community formation. *Convergence*, 0(0).

Attwood, F.; Hakim, J.; and Winch, A. 2017. Mediated intimacies: Bodies, technologies and relationships.

Barbosa, N. M.; Wang, G.; Ur, B.; and Wang, Y. 2021. Who Am I?: A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3): 1–32.

Barocas, S.; and Selbst, A. D. 2016. Big Data's Disparate Impact. *California Law Review*, 104(3): 671–732.

Battisti, D. 2025. Second-Person Authenticity and the Mediating Role of AI: A Moral Challenge for Human-to-Human Relationships? *Philosophy & Technology*, 38(1): 28.

Bennett, L. 2013. Researching Online Fandom. *Cinema Journal*, 52(4): 129–134.

Best, P.; Manktelow, R.; and Taylor, B. 2014. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review*, 41: 27–36.

Bhat, M. 2025. How Dynamic vs. Static Presentation Shapes User Perception and Emotional Connection to Text-Based AI. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, IUI '25, 846–860. New York, NY, USA: Association for Computing Machinery. ISBN 9798400713064.

Bhat, M.; and Long, D. 2024. Designing Interactive Explainable AI Tools for Algorithmic Literacy and Transparency. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, 939–957. New York, NY, USA: Association for Computing Machinery. ISBN 9798400705830.

Binns, R.; Veale, M.; Van Kleek, M.; and Shadbolt, N. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

Blut, M.; Wang, C.; Wünderlich, N. V.; and Brock, C. 2021. Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 49: 632–658.

Borau, S. 2024. Deception, discrimination, and objectification: Ethical issues of female AI agents. *Journal of Business Ethics*, 1–19.

Bozdağ, A. A. 2024. The AI-mediated intimacy economy: a paradigm shift in digital interactions. *AI & Society*.

Braun, V.; and Clarke, V. 2022. Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. *International Journal of Transgender Health*, 24(1): 1–6.

Brenncke, M. 2024. Regulating dark patterns. *Notre Dame J. Int'l Comp. L.*, 14: 39.

Brooks, R. 2021. *Artificial intimacy: Virtual friends, digital lovers, and algorithmic matchmakers*. Columbia University Press.

Bury, R. 2016. Technology, fandom and community in the second media age. *Convergence*, 23(6): 627–642.

Chang, F.; and Herath, D. 2025. From Interaction to Relationship: The Role of Memory, Learning, and Emotional Intelligence in AI-Embodied Human Engagement. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 1269–1273. IEEE.

Chaturvedi, R.; Verma, S.; Das, R.; and Dwivedi, Y. K. 2023. Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting and Social Change*, 196: 122634.

Chromik, M.; Eiband, M.; Völkel, S. T.; and Buschek, D. 2019. Dark Patterns of Explainability, Transparency, and User Control for Intelligent Systems. In *Joint Proceedings of the ACM IUI 2019 Workshops*, 1–6. Los Angeles, USA: ACM.

Cockayne, D.; Leszczynski, A.; and Zook, M. 2017. # HotForBots: Sex, the non-human and digitally mediated spaces of intimate encounter. *Environment and Planning D: Society and Space*, 35(6): 1115–1133.

Crawford, K. 2021. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Crowder, J. A.; and Friess, S. 2011. Metacognition and Metamemory Concepts for AI Systems. In *Proceedings of the International Conference on Artificial Intelligence (ICAI)*, 1–6. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Data.ai. 2024. Top Mobile Apps Worldwide for March 2024 by Downloads.

De Freitas, J.; and Cohen, I. G. 2024. The health risks of generative AI-based wellness apps. *Nature Medicine*, 30: 1269–1275.

De Freitas, J.; Uguralp, A. K.; Uguralp, Z. O.; and Puntoni, S. 2024. AI Companions Reduce Loneliness. *arXiv preprint arXiv:2407.19096*.

Djufril, R.; Frampton, J. R.; and Knobloch-Westerwick, S. 2025. Love, Marriage, Pregnancy: Commitment Processes in Romantic Relationships with AI Chatbots. *Computers in Human Behavior: Artificial Humans*, 100155.

Döring, N.; Le, T. D.; Vowels, L. M.; Vowels, M. J.; and Marcantonio, T. L. 2025. The Impact of Artificial Intelligence on Human Sexuality: A Five-Year Literature Review 2020–2024. *Current Sexual Health Reports*, 17(1): 1–39.

Erikson, E. H. 1968. *Identity: Youth and Crisis*. W. W. Norton & Company.

Fast, E.; and Horvitz, E. 2021. How Users Talk About Chatbots: Emotional Reactions, Mental Models, and Expectations. *Journal of Human-Computer Interaction*, 37(2): 180–197.

Forattini, F. M.; Connolly, R.; and Joshi, K. 2025. Breaking Boundaries: Advancing Gender and Technology Research to Combat Sextortion. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 56(1): 6–10.

Fosch Villaronga, E. 2019. "I Love You," Said the Robot: Boundaries of the Use of Emotions in Human-Robot Interactions. *Emotional Design in Human-Robot Interaction: Theory, Methods and Applications*, 93–110.

Ge, L.; and Hu, T. 2025. Gamifying intimacy: AI-driven affective engagement and human-virtual human relationships. *Media, Culture & Society*, 01634437251337239.

George, A. S.; George, A. S. H.; Baskar, T.; and Pandey, D. 2024. The Allure of Artificial Intimacy: Examining the Appeal and Ethics of Using Generative AI for Simulated Relationships. *Zenodo*.

Giles, D. C. 2002. Parasocial Interaction: A Review of the Literature and a Model for Future Research. *Media Psychology*, 4(3): 279–305.

Gordon-Tapiero, A. 2025. A Liability Framework for AI Companions. *George Washington Journal of Law & Technology*. Forthcoming.

Grubbs, J. B.; Perry, S. L.; Wilt, J. A.; and Reid, R. C. 2019. Pornography Problems Due to Moral Incongruence: An Integrative Model with a Systematic Review and Meta-Analysis. *Archives of Sexual Behavior*, 48(2): 397–415.

Hancock, J. T.; Naaman, M.; and Levy, O. 2020. Artificial Intelligence, Trust, and Communication in the Age of Synthetic Media. *Journal of Computer-Mediated Communication*, 25(1): 100–109.

He, J. K.; Wallis, F. P.; Gvirtz, A.; and Rathje, S. 2024. Artificial intelligence chatbots mimic human collective behaviour. *British Journal of Psychology*.

Ho, M.-T.; Mantello, P.; and Vuong, Q.-H. 2024. Emotional AI in education and toys: Investigating moral risk awareness in the acceptance of AI technologies from a cross-sectional survey of the Japanese population. *Heliyon*, 10(16): e36251.

Hoskins, A. 2024. AI and Memory. *AI and Memory Collection*. Published online 11 September 2024.

Hu, B.; Mao, Y.; and Kim, K. J. 2023. How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception. *Computers in Human Behavior*, 145: 107760.

Hu, R. 2019. A Chatbot, a Therapist, or a Friend: A Study on the Sociality of Woebot and the Social Acceptability of it.

Jenkins, H. 1992. *Textual Poachers: Television Fans and Participatory Culture*. Routledge.

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

Jokinen, K.; and Wilcock, G. 2023. Do You Remember Me? Ethical Issues in Long-term Social Robot Interactions. In *Proceedings of the IEEE Conference on Human-Robot Interaction*. IEEE.

Jones, M. L.; Kaufman, E.; and Edenberg, E. 2018. AI and the Ethics of Automating Consent. *IEEE Security & Privacy*, 16(3): 64–72.

Kang, E. J.; Kim, H.; Kim, H.; Fussell, S. R.; and Kim, J. 2025. Can Fans Build Parasocial Relationships through Idols' Simulated Voice Messages?: A Study of AI Private Call Users' Perceptions, Cognitions, and Behaviors. *Proceedings of the ACM on Human-Computer Interaction*, 9(2): CSCW044:1–31.

Khampuong, P.; Nilsook, P.; and Wannapiroon, P. 2023. Artificial Intelligence Avatar for Conversational Agent. In *2023 IEEE Conference Proceedings*. IEEE.

Kirk, H. R.; Vidgen, B.; Röttger, P.; and Hale, S. A. 2024. The benefits, risks and bounds of personalizing the alignment of large language models to individuals. *Nature Machine Intelligence*, 6(4): 383–392.

Kouros, T.; and Papa, V. 2024. Digital Mirrors: AI Companions and the Self. *Societies*, 14(10): 200.

Kroczek, L. O. H.; May, A.; Hettenkofer, S.; Ruider, A.; Ludwig, B.; and Mühlberger, A. 2024. The Influence of Persona and Conversational Task on Social Interactions with a LLM-Controlled Embodied Conversational Agent. *arXiv preprint arXiv:2411.05653*.

Lamerichs, N. 2011. Stranger than fiction: Fan identity in cosplay. *Transformative Works and Cultures*, 7.

Lee, E. 2024. Towards Ethical Personal AI Applications: Practical Considerations for AI Assistants with Long-Term Memory. *arXiv preprint arXiv:2409.11192*.

Li, E. C.-Y.; and Pang, K.-W. 2024. Fandom meets artificial intelligence: Rethinking participatory culture as human–community–machine interactions. *European Journal of Cultural Studies*, 27(4): 778–787.

Liao, Q. V.; and Sundar, S. S. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*, 18 pages. New York, NY, USA: Association for Computing Machinery.

Liu, B.; Kang, J.; and Wei, L. 2024. Artificial intelligence and perceived effort in relationship maintenance: Effects on relationship satisfaction and uncertainty. *Journal of Social and Personal Relationships*, 41(5): 1232–1252.

Livingstone, S.; and Bulger, M. 2014. Children's rights in the digital age: A download from children around the world. *UNICEF Office of Research*.

Locatelli, C.; et al. 2018. *" The Perfect Companion": From Cyborgs to Gynoids-Sex Robots and the Commodification of Authentic Intimate Experience*. Master's thesis, MASTER ERASMUS MUNDUS EN ESTUDIOS DE LAS MUJERES Y DE GÉNERO.

Lorenz, T. 2023. An influencer's AI clone will be your girlfriend for $1 a minute. *The Washington Post*.

Loveys, K.; Hiko, C.; Sagar, M.; Zhang, X.; and Broadbent, E. 2021. "I felt her company": A qualitative study on factors affecting closeness and emotional support seeking with an embodied conversational agent. *International Journal of Human-Computer Studies*, 155: 102771.

Luger, E.; and Sellen, A. 2016. Like Having a Really Bad PA: The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297. ACM.

Luxton, D. D. 2014. Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine*, 62(1): 1–10.

Lühring, J.; Shetty, A.; Koschmieder, C.; Garcia, D.; Waldherr, A.; and Metzler, H. 2024. Emotions in Misinformation Studies: Distinguishing Affective State from Emotional Response and Misinformation Recognition from Acceptance. *Cognitive Research: Principles and Implications*, 9(82).

Maeda, T.; and Quan-Haase, A. 2024. When Human-AI Interactions Become Parasocial: Agency and Anthropomorphism in Affective Design. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 1068–1077. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Maeng, W.; and Lee, J. 2022. Designing and evaluating a chatbot for survivors of image-based sexual abuse. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, 1–21.

Mahajan, P. 2025. Beyond Biology: AI as Family and the Future of Human Bonds and Relationships.

Marinucci, L.; Mazzuca, C.; and Gangemi, A. 2023. Exposing implicit biases and stereotypes in human and artificial intelligence: state of the art and challenges with a focus on gender. *AI & SOCIETY*, 38(2): 747–761.

Markelius, A.; Wright, C.; Kuiper, J.; et al. 2024. The mechanisms of AI hype and its planetary and social costs. *AI Ethics*, 4: 727–742.

Maslych, M.; Pumarada, C.; Ghasemaghaei, A.; and Jr, J. J. L. 2025. Takeaways from Applying LLM Capabilities to Multiple Conversational Avatars in a VR Pilot Study. *arXiv preprint arXiv:2501.00168*.

Mason-Bertrand, A. 2019. *Cosplay: An Ethnographic Study of Subculture and Escape*. Phd thesis, University of Sheffield.

Merrill Jr, K.; Kim, J.; and Collins, C. 2022. AI companions for lonely individuals and the role of social presence. *Communication Research Reports*, 39(2): 93–103.

Morris, R. R.; Kouddous, K.; Kshirsagar, R.; and Schueller, S. M. 2021. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. *Journal of Medical Internet Research*, 23(5): e16115.

Noor, N.; Hill, S. R.; and Troshani, I. 2021. Artificial Intelligence Service Agents: Role of Parasocial Relationship. *Journal of Computer Information Systems*, 62(6): 1009–1023.

Nowotny, H. 2024. AI and the illusion of control. In *Proceedings of the Paris Institute for Advanced Study*, volume 1.

Obiefuna, P. 2025. Relational Presence in AI: Architecture, Vulnerabilities, and Stewardship. *Vulnerabilities, and Stewardship (April 29, 2025)*.

Oni, O. 2024. Memory-Enhanced Conversational AI: A Generative Approach for Context-Aware and Personalized Chatbots. *Computational and Processing Systems*, 12(2): 123–139.

Pataranutaporn, P.; Danry, V.; Leong, J.; Punpongsanon, P.; Novy, D.; Maes, P.; and Sra, M. 2021. AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence*, 3(12): 1013–1022.

Pater, J.; McDaniel, B. T.; Farhat Nova, F.; Drouin, M.; O'Connor, K.; and Zytko, D. 2025. A Commentary on Sexting, Sextortion, and Generative AI: Risks, Deception, and Digital Vulnerability. *Family Relations*. First published: 19 February 2025.

Pfadenhauer, M. 2015. The contemporary appeal of artificial companions: Social robots as vehicles to cultural worlds of experience. *The Information Society*, 31(3): 284–293.

Possati, L. M. 2023. Psychoanalyzing artificial intelligence: the case of Replika. *AI & SOCIETY*, 38: 1725–1738.

Prochaska, J. J.; Vogel, E. A.; Chieng, A.; Kendra, M.; Baiocchi, M.; Pajarito, S.; and Robinson, A. 2021. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. *Journal of medical Internet research*, 23(3): e24850.

Ragab, A.; Mannan, M.; and Youssef, A. 2024. "Trust Me Over My Privacy Policy": Privacy Discrepancies in Romantic AI Chatbot Apps. In *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 484–495. IEEE.

Reineke, M. J. 2022. The Touching Test: AI and the Future of Human Intimacy. *Contagion: Journal of Violence, Mimesis, and Culture*, 29: 123–146.

Rupprecht, T.; Chang, S.-E.; Wu, Y.; Lu, L.; Nan, E.; hsiang Li, C.; Lai, C.; Li, Z.; Hu, Z.; He, Y.; Kaeli, D.; and Wang, Y. 2024. Digital Avatars: Framework Development and Their Evaluation. *arXiv preprint arXiv:2408.04068*.

Schwartz, M. F.; and Southern, S. 2017. Recovery from sexual compulsivity. *Sexual Addiction & Compulsivity*, 24(3): 224–240.

Stevenson, A. C. 2022. *A Posthuman Analysis: Cosplay Identity Creation, Investment and 'Becomings' through Social Media Usage*. Ph.D. thesis, Leeds Trinity University.

Subrahmanyam, K.; Smahel, D.; and Greenfield, P. 2001. The Internet and identity development in adolescence. *Applied Developmental Psychology*, 22(1): 7–30.

Subrahmanyam, K.; and Šmahel, D. 2010. Constructing Identity Online: Identity Exploration and Self-Presentation. In *Digital Youth: Advancing Responsible Adolescent Development*, 59–80. Springer.

Sutcliffe, B. 2024. *Artificial Intimacy: Exploring Intimacy in Human and AI-enabled Chatbot Relations: Its Existence, Its Authenticity and Its Moral Implications*. Ph.D. thesis.

Thingbø Mlonyeni, P. M. 2025. Personal AI, deception, and the problem of emotional bubbles. *AI & SOCIETY*, 40: 1927–1938.

Tosenberger, C. 2008. Homosexuality at the online Hogwarts: Harry Potter slash fanfiction. *Children's Literature*, 36: 185–207.

Turkle, S. 2017. Empathy Machines: Forgetting the Body. In Barrett, J. P., ed., *A Psychoanalytic Exploration of the Body in Today's World*, 11–22. Routledge. ISBN 9781315159683. First published in 2017, 1st edition.

Ventura, A.; Starke, C.; Righetti, F.; and Köbis, N. 2025. Relationships in the Age of AI: A Review on the Opportunities and Risks of Synthetic Relationships to Reduce Loneliness.

Waytz, A.; Cacioppo, J.; and Epley, N. 2014. Anthropomorphizing technology: When robots, computers, and other agents become social actors. *Current Directions in Psychological Science*, 23(6): 394–398.

Weigel, S.; and Rudnick, J. 2022. The Use and Importance of Gaming and Roleplay in Identity Negotiation. *Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato*.

Wängqvist, M.; and Frisén, A. 2016. Who am I online? Understanding the meaning of online contexts for identity development. *Adolescent Research Review*, 1(2): 139–151.

Xu, C.; Li, P.; Wang, W.; Yang, H.; Wang, S.; and Xiao, C. 2022. COSPLAY: Concept Set Guided Personalized Dialogue Generation Across Both Party Personas. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 201–211. ACM.

Zheng, D. 2022. The empowering role of online fan communities for LGBTQ+ youth. *New Media & Society*.