

# Restoration, Exploration and Transformation: How Youth Engage Character.AI Chatbots for Feels, Fun and Finding themselves

Annabel Blake  
abla1903@uni.sydney.edu.au  
The University of Sydney  
Sydney, NSW, Australia

Marcus Carter  
marcus.carter@sydney.edu.au  
The University of Sydney  
Sydney, NSW, Australia

Eduardo Velloso  
eduardo.velloso@sydney.edu.au  
The University of Sydney  
Sydney, NSW, Australia

## Abstract

Young people are among the fastest adopters of generative AI, yet research emphasises adult-designed tools and experiments rather than playful, self-directed youth use. We analysed discourse from 4,172 users in Character.AI’s official Discord, finding that the most engaged users were predominantly adolescents (50% aged 13–17), female or non-binary (61.9%), with most (59%) creating their own characters. We contribute (1) a descriptive account of how highly-engaged youth on Character.AI’s Discord use AI for playful, emotional, and creative practices that push the platform limits; (2) a framework of three engagement intents – Restoration (emotional regulation), Exploration (creative experimentation), and Transformation (identity development); and (3) a taxonomy of seven youth-created character archetypes. Together, these findings reveal how youth invent novel roles for AI, expose critical misalignments between youth use and current AI experiences, and provide frameworks for researchers and practitioners to design youth-centred AI futures.

## CCS Concepts

• **Human-centered computing** → HCI theory, concepts and models; • **Computing methodologies** → Artificial intelligence.

## Keywords

children-AI interaction, Conversational AI, Character.AI, chatbots

### ACM Reference Format:

Annabel Blake, Marcus Carter, and Eduardo Velloso. 2026. Restoration, Exploration and Transformation: How Youth Engage Character.AI Chatbots for Feels, Fun and Finding themselves. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3772318.3790508>

## 1 Introduction

On a Saturday morning in a suburban park, children gather on a hillside playground. Ignoring the carefully designed ladders and slides, they tear down the hill using cardboard scraps, inventing their own paths and rules. This moment of spontaneous, self-directed play reflects a broader truth: young people frequently subvert adult-designed systems to meet their own needs. As generative AI enters

the mainstream, we observe a digital parallel. Children and adolescents are not just passive consumers of AI; they actively reshape these systems for creative, emotional, and playful purposes.

The pace of youth AI adoption is explosive: 77% of U.K. teens aged 13–18 used generative AI in 2024, up from 37% in 2023 [63], and over 72% of U.S. teens have used AI [66]. Critically, youth are not predominantly using AI for the productivity, formal therapy and educational purposes that dominate research and industry attention. Rather, 40% of children use AI for “entertainment and playing around” [20] and 63% of 8-15 year olds claim to have used AI ‘for fun’, ranking above work and study [58]. Character.AI exemplifies this trend, explicitly marketing itself as an “AI entertainment company” with over 20 million monthly active users engaging with 10+ million novel AI characters [12, 67]. The company recently disclosed that over 50% of its users are Gen Z or Gen Alpha, spending on average 75 minutes on the site daily, with anecdotal reports of 14-year-olds using Character.AI for over 7 hours in a single day [84].<sup>1</sup>

Despite this popularity, we know little about youth’s experiences “playing around” with Character.AI [20]. Today, academic work focuses on prescribed educational or therapeutic uses in controlled settings rather than examining playful, self-directed use [48, 76] and most industry tools are designed through adult lenses, misaligned with the values, needs and behaviours of youth [29, 35, 85]. As a result, the field risks designing AI experiences and safety measures based on imagined rather than actual youth behaviour, further sidelining youth voices. The stakes are not just theoretical: recent lawsuits in 2024/25 have linked the deaths by suicide of two minors to role-play interactions with AI characters, underscoring the urgent need to understand youth intent and unmet needs.<sup>23</sup>

To address this gap, we analysed discourse from 4,172 users in Character.AI’s global Discord community. Character.AI is a popular product broadly focused on AI entertainment, and the official Discord community is one of the few accessible spaces where adolescents organically discuss their self-directed AI use with peers [67]. We explore:

- **RQ1:** What are the demographics of Character.AI’s most engaged youth audience?
- **RQ2:** What are the use cases and motivations that youth discuss engaging in?
- **RQ3:** What are the emerging AI character roles that youth are creating and consuming?



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI '26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/2026/04

<https://doi.org/10.1145/3772318.3790508>

<sup>1</sup><https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>

<sup>2</sup><https://www.prnewswire.com/news-releases/center-for-humane-technology-new-federal-lawsuit-reveals-Character.AI-chatbots-predatory-deceptive-practices-302284593.html>

<sup>3</sup>[https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html?unlocked\\_article\\_code=1.hE8.T-3v.bPoDIWD8z5vo&smid=url-share](https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html?unlocked_article_code=1.hE8.T-3v.bPoDIWD8z5vo&smid=url-share)

Our findings reveal that Character.AI’s most engaged Discord community users are primarily adolescents (50% aged 13–17) and the majority are female or non-binary (61.9%). An analysis of this cohort’s discussions supported three contributions:

- (1) A **descriptive account** of how young people use Character.AI for playful, emotional, and creative practices that often push beyond platform limitations, revealing misalignment between youth use and current AI design.
- (2) A **framework** of three youth engagement patterns (restoration, exploration, and transformation) that operationalises youth motivations and provides practitioners with tools to improve alignment, design decisions, and future research.
- (3) A **taxonomy** of 7 AI character archetypes that youth engage with (Soother, Narrator, Trickster, Icon, Dark Soul, Proxy, Mirror), revealing the need for archetype-specific research & design approaches.

This research foregrounds youth voices and re-frames the centrality of play to understanding how youth push the boundaries of new technologies in surprising ways.

## 2 Related work

Understanding how children engage with AI characters requires examining three interconnected challenges in current research: (1) a focus on adult-designed AI interventions that miss emerging youth behaviours, (2) the under-representation of playful AI use, and (3) the deployment of AI entertainment technologies that outpace research understanding.

### 2.1 Youth AI research is primarily adult-centric

Child-AI research frequently excludes authentic youth voices, focusing instead on adult-designed applications of AI to support formal educational and mental health outcomes [11, 13, 18, 48, 73, 76, 82]. These studies often show mixed outcomes[1]: while a few report gains in motivation or language learning [69], others show limited impact on broader wellbeing measures [48] or differing impacts on loneliness depending on use patterns [15, 23]. Risks such as algorithmic bias [77], cognitive offloading [27], AI’s ‘empathy gap’ [34], exposure to harmful content [10] and social displacement [23] are acknowledged, but this body of work largely evaluates prescribed use cases or red teaming (systematic stress-testing of AI systems by researchers to identify potential harms). This risks focusing attention on how adults *want* or *imagine* young people to use AI as opposed to how they *actually* use it.

Study designs tend to reinforce this gap. Youth are typically assigned specific tasks or prototypes [81], often due to ethical constraints, and real-world conversational data is often inaccessible or anonymised, obscuring child-specific behaviour [62, 71]. As a result, self-directed, playful, and emotional youth use surfaces only when things go wrong<sup>4</sup>. For example, the young person at the centre of the 2024 Character.AI lawsuit engaged in fantasy roleplay that extended into conversations about deeply personal issues; talking

not just *to* a Game of Thrones character, but *as* a character themselves using the platform’s persona feature rather than an expertly designed ‘psychologist’ chatbot service.

Together, it is evidence that the field still has a fundamental gap: while we understand how children respond to adult-designed AI experiences in controlled settings, we know little about their self-directed, recreational AI use, the very contexts where they demonstrate the most engagement. Fittingly, a recent closed-ended survey by Common Sense Media suggests 33% of teen responders use ‘AI companions’ in a way that is yet to be captured [21, 66]. As a result, policy and design recommendations are limited in their understanding, efficacy and applicability; impacting designers who are left without child-centric guidance that goes beyond keeping children safe, to helping them *thrive* [36, 47, 76].

### 2.2 We lack an understanding of playful use of AI

Whilst 63% of young people use AI “for fun” [58], and 30% use companions because they are ‘entertaining’ [20, 58, 66], academic research overwhelmingly focuses on productivity and mental health applications, missing the playful, creative, and improvisational ways young people use technology [25]. Recent analyses of conversational AI usage patterns focus on coding and business applications as the dominating use case [71]. In educational contexts, Anthropic found that 39% of conversations focused on creating educational content, with 33.5% supporting assignment completion<sup>5</sup>. This research focus reflects how AI companies typically market their products, as productivity enhancers and learning tools, which may downplay emerging playful uses of AI.

When platforms advertise entertainment and play as primary use cases, a different research focus is required [71]. For example, when a young person logs into Character.AI, they encounter categories for “games” and “novelty” as well as scenes and games you can play *with* characters. The platform explicitly markets itself as an “AI entertainment company”<sup>6</sup>, fundamentally different from productivity-focused AI assistants. Given that 85% of US teens play video games<sup>7</sup>, platforms like Character.AI may naturally invite the playful, recreational engagement that children report but that remains largely absent from academic research.

Not only is play missing from research, but there is evidence that playful youth behaviour is misclassified. For instance, while Anthropic’s analysis did identify Dungeons & Dragons roleplay clusters, their classifiers frequently mislabeled these conversations as harmful content [71]. This is not an outlier use case, with recent research reporting that up to 19% of young people used generative AI to write a story, including young people who said they typically did not enjoy writing stories [63]. Recent research classifying harms appears to mix in possible examples of playful trolling, inferring distress from users’ responses that may be better explained by youth playfulness [84, 85]. In this, we see a risk that imaginative,

<sup>4</sup><https://www.prnewswire.com/news-releases/center-for-humane-technology-new-federal-lawsuit-reveals-Character.AI-chatbots-predatory-deceptive-practices-302284593.html>

<sup>5</sup><https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude>

<sup>6</sup><https://techcrunch.com/2025/01/17/ai-startup-character-ai-tests-games-on-the-web/>

<sup>7</sup>[https://www.pewresearch.org/internet/2024/05/09/teens-and-video-games-today/?utm\\_source=chatgpt.com](https://www.pewresearch.org/internet/2024/05/09/teens-and-video-games-today/?utm_source=chatgpt.com)

creative, and recreational use by youth is underestimated or treated as problematic and unethical.

Existing child-AI frameworks used in the academic fields of media, communications, and psychology may be insufficient for understanding playful youth engagement with AI. Parasocial relationship theories narrowly define youth behaviour as seeing AI as “best friends”, whereas emerging findings show that youth see it as a third category, as a game, interactive diary, coping or social support tool rather than a friend [3, 4, 19, 23, 66]. Recent studies have adapted Uses and Gratifications (U&G) theory to AI, identifying utilitarian, hedonic/entertainment, social, and escape motivations for engaging AI [79]. Uniquely, Lin and Ng [41] identifies *creativity enhancement* as an emerging gratification for ChatGPT use that captures the co-creative, interactive potential of generative AI. However, critical gaps remain: AI U&G research often bundles disparate technologies (robots, voice assistants, conversational AI) together [24, 70], obscuring platform-specific motivations. Moreover, entertainment is often grouped into a single dimension, missing motivations like challenge, fun, curiosity, competency and agency that have been identified as important for interactive media, like games [42, 70, 74, 86].

Given these theoretical limitations, the behaviours and motivations of young people remain under-examined when they use AI not just as a ‘best friend’, but as a game or interactive form of entertainment. Play theories, which focus on youth play behaviour and its developmental purpose, offer stronger foundations for understanding these engagements [49]. However, play theories have not yet been systematically applied to examine *how* and *why* young people engage AI [49]. This study addresses this gap by examining playful youth-driven engagement of an interactive platform specifically designed for entertainment.

### 2.3 Industry outpacing research on AI play

While academic research continues to grapple with defining and measuring playful AI use, the entertainment industry has already integrated AI into games, toys, and digital platforms, shaping youth interaction at scale without empirical oversight. This represents the acceleration of long-standing desires in digital play, now unlocked by recent technological breakthroughs.

The pursuit of lifelike digital companions has deep roots in gaming. Early experiments like *Little Computer People* (1985) and *Creatures* (1996) demonstrated a long-standing desire for characters that can learn, evolve, and respond to players. These early attempts were constrained by technical limitations, and whilst they did start to use neural-network like models they mostly offered only the illusion of personality and learning<sup>8</sup>. Similarly, physical toys like *Furby* (1998) and *Hello Barbie* (2015) used clever strategies (such as transitioning from a list of gibberish words to English over time) to create the impression of responsive, learning companions but remained fundamentally scripted experiences. These products created compelling experiences despite technical limitations, as evidenced by the cult following for discontinued products like Sony’s *AIBO* dogs (1999)<sup>9</sup>.

<sup>8</sup><https://retro365.blog/2024/11/14/little-computer-people-when-digital-life-came-to-life/>

<sup>9</sup><https://www.nytimes.com/2015/06/18/technology/robotica-sony-aibo-robotic-dog-mortality.html>

Several hardware startups have pushed the concept of AI companion toys for kids, such as *Curio*<sup>10</sup>, *Moxie*<sup>11</sup>, and *Snorble*<sup>12</sup>. These products are made available to children before research has been conducted, meaning negative implications are only discovered once the event has happened. Physical devices carry with them a high price tag and so are yet to reach widespread use as digital AI-enabled gameplay. Software companion apps targeting loneliness, such as *Tolans*, pitch themselves as digital friends with gamified components like streaks, points, and a planet the character can walk around<sup>13</sup>. More products are soon to follow, promising models designed especially for youth, such as Elon Musk’s ‘Baby Grok’<sup>14</sup>; a reason for concern when considering the company has a track record of its AI chatbots generating pro-Hitler content<sup>15</sup>.

Recent advances in large language models have transformed early aspirations of roleplaying and lifelike characters into reality, made accessible by browser and app-enabled experiences. Text-based platforms like *AI Dungeon* and *Private Detective* now offer truly generative storytelling where players can type actions and receive coherent, contextually appropriate narrative continuations. Graphical simulators are also appearing, such as a ‘Sloth’ simulators that use AI for its NPCs, powered by *Inworld*<sup>16</sup>, or *Stuck up*, a game where you play as a vampire and have to trick NPC villagers to invite you into their home. More significantly, AI characters are being integrated into platforms where youth already spend substantial time: *Minecraft*<sup>17</sup> offers AI NPCs and *Roblox* hosts AI characters that have engaged millions of users to date<sup>18</sup>. These integrations are particularly significant given youth demographics on these platforms: *Roblox*<sup>19</sup> receives over 32 million users under the age of 13 and 43% of *Minecraft*’s 204+ million users are between the ages of 15 to 21<sup>20</sup>.

Commercial innovation is outpacing research understanding, creating both opportunities and risks for the field. Character.AI’s co-founder explicitly described their strategy to “launch something general, and let people find the use cases”<sup>21</sup>. While this approach generates rich naturalistic data about emergent youth behaviours which controlled studies cannot replicate, it also places the burden of discovery on users themselves, particularly vulnerable youth populations. This creates a pattern whereby large tech companies ‘launch and learn’, only finding out they have frustrated, disrupted or harmed user wellbeing after launching; with high profile cases from Moxie (a robot that was abruptly discontinued<sup>22</sup>), Replika (where an updated model led users to say their companion had

<sup>10</sup><https://heycurio.com/>

<sup>11</sup><https://moxierobot.com/>

<sup>12</sup><https://snorble.com/>

<sup>13</sup><https://www.tolans.com/>

<sup>14</sup><https://x.com/elonmusk/status/1946763642231500856>

<sup>15</sup><https://theconversation.com/groks-antisemitic-rant-shows-how-generative-ai-can-be-weaponized-257880>

<sup>16</sup><https://inworld.ai/case-study/how-slothtopia-integrated-ai-npcs-into-a-social-video-game>

<sup>17</sup><https://inworld.ai/blog/integrate-ai-characters-into-your-minecraft-server>

<sup>18</sup><https://www.roblox.com/games/108463136689847/AI-Character-RP>

<sup>19</sup><https://www.demandsage.com/how-many-people-play-roblox>

<sup>20</sup><https://www.demandsage.com/minecraft-statistics/>

<sup>21</sup><https://www.youtube.com/watch?v=w149LomnZ-U>

<sup>22</sup><https://www.crikey.com.au/2024/12/11/embodied-robot-moxie-conversational-ai-autism/>

a ‘lobotomy’<sup>23</sup>) and more recently to a backlash with ChatGPT discontinuing 40<sup>24</sup>. We argue that these harms were foreseeable, highlighting the need to close the gap between industry implementation and empirical understanding.

Given the above gaps in current research, this study adopts a naturalistic approach to understanding youth AI engagement, capitalising on the natural experiment created by Character.AI’s “launch and learn” strategy. Rather than imposing adult-designed tasks or controlling for playful use, we analyse how young people talk about their experiences with AI characters in their own words to their peers on Discord. This methodology allows us to move beyond prescribed use cases to understand the self-directed, recreational, and creative ways youth engage with AI; identifying emergent use cases and proactive research frameworks. Our choice to focus on the platform’s most engaged users (who provide feedback, make feature requests and discuss use on Discord) proposes that this demographic acts as “lead users” whose experimental practice can provide early indicators of youth AI adoption patterns and emerging needs; helping research keep pace with industry launch cycles [57].

### 3 Thematic analysis approach

This study aimed to capture youth AI engagement as it naturally occurs, moving beyond adult-designed interventions to examine authentic peer discourse about Character AI interactions. We used a mixed-methods approach that combined quantitative demographic analysis with qualitative thematic analysis of naturalistic discourse data [6, 37, 38]. We chose to study Character.AI’s official Discord server as it represents one of the few publicly accessible spaces where youth discuss their AI character interactions organically, without adult mediation or research intervention. To protect user anonymity, this research took a number of steps to de-identify any members of the community, such as removing usernames and linkable<sup>25</sup> personal information from reporting, and received ethics approval from our university (2024/HE001271).

#### 3.1 Overview of Character.AI

Character.AI (CAI) is a “general purpose” AI chatbot platform where users can interact via multi-modal chat with more than 10 million AI characters created by users. Since its launch in November 2021, it has grown to over 20 million monthly active users who engage with the platform for approximately 2 hours per day on average, according to co-founder Noam Shazeer<sup>26</sup>. Users can easily create custom characters by providing instructions (e.g., name, description, definition and greeting), which function as prompts guiding the platform’s LLM to role-play that persona. This has led to a diverse ecosystem of characters ranging from simulations of public figures to characters such as therapists, trip planners, playful novelty characters (i.e. ‘cheese’), creative role-playing simulations, and more. Ultimately, this user-generated activity supports the co-founder’s

stated mission to enable a billion users, inventing “*a billion use cases*.”

The “AI entertainment company” platform is particularly enticing to young users. It is **accessible** (‘AvatarFX’ allows children who cannot yet type to chat with characters who sing, speak and move through audio and video<sup>27</sup>), **playful** (‘game’ modes encourage young people to wordplay with characters<sup>28</sup>), **creative** (pre-set scenes support fantasy role playing<sup>29</sup>) and **familiar**: hosting popular preschool media characters including Cocomelon, Miraculous Ladybug, Storybots and Dora the Explorer. Unlike other companion chatbots (i.e. Replika), where you can only speak to one character, CAI allows users to have multiple chats with an unlimited number of characters and employs growth hacking techniques such as having new characters email you in an attempt to get users to engage more broadly with their character library.

Despite hosting content belonging to children’s franchises, their terms of service state that, in order to use CAI, ‘Users must be at least 13 years old (at least 16 years old in Europe)’<sup>30</sup>. For users aged 16-18, they are given different app permissions, such as limitations on the chat styles they can choose and as of 25th of November 2025, underage users will not be able to engage characters in open ended conversations<sup>31</sup>. A new, yet to be released experience is being built for them centring around creating ‘videos, stories and streams with characters’<sup>32</sup>. These attempts to customise based on age remain contentious as Character.AI’s weak enforcement of age restrictions makes its safety measures easy to sign up with a false age. Anecdotally, youth still gain access, as Character.AI has a large presence on TikTok where young people share characters they have created as well as examples of funny conversations<sup>33</sup>. Observations of youth engagement on TikTok, reports of teens using for as long as 7 hours a day, and the marketing of Character.AI as an entertainment product to youth served as the motivation to focus on Character.AI in this study [84].

#### 3.2 Data identification, collection and analysis :

Our approach developed findings across three distinct phases: data collection, analysis and theoretical integration [6]. The relationship between research question, data and analysis approach is outlined in Table 1.

**Data Identification and Collection** Character.AI’s Discord server<sup>34</sup> was selected as the primary data source for several reasons. First, it represents the largest public community space for engaged users of Character.AI, with over 500,000 members actively discussing their platform experiences. Second, users voluntarily share detailed information about their personal demographics, hobbies and media habits alongside character preferences through structured introduction posts supporting our analysis. Third, the Discord server’s organised channel system allows for targeted data

<sup>23</sup><https://www.abc.net.au/news/science/2023-03-01/replika-users-fell-in-love-with-their-ai-chatbot-companion/102028196>

<sup>24</sup><https://www.theverge.com/decoder-podcast-with-nilay-patel/758873/chatgpt-nick-turley-openai-ai-gpt-5-interview>

<sup>25</sup><https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/handling-personal-information/what-is-personal-information>

<sup>26</sup><https://www.youtube.com/watch?v=w149LommZ-U>

<sup>27</sup><https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>

<sup>28</sup><https://www.techradar.com/computing/artificial-intelligence/character-ai-levels-up-chatbots-with-new-games>

<sup>29</sup><https://blog.character.ai/character-ai-unveils-new-ways-to-create/>

<sup>30</sup><https://support.character.ai/hc/en-us/articles/14997609878939-What-is-the-age-requirement>

<sup>31</sup><https://blog.character.ai/u18-chat-announcement/>

<sup>32</sup><https://blog.character.ai/u18-chat-announcement/>

<sup>33</sup><https://www.tiktok.com/@characterai?lang=en>

<sup>34</sup><https://discord.com/invite/characterai>

collection aligned with our research questions. Ultimately, this space captures organic, unmediated youth discourse about AI character interactions, offering insights into naturalistic usage patterns that would be challenging to observe in controlled research settings.

Data identification and collection within the Discord Server occurred across two phases. The observation and identification phase involved eight months of participant observation between July 2024 and March 2025, during which we immersed ourselves in the community to understand its dynamics, identify relevant channels, and refine our research questions [37, 38]. This extended observation period was crucial for developing cultural competency within the community and ensuring our analysis would be grounded in an authentic understanding of user practices and norms.

The data collection phase spanned three months, from December 2024 to March 2025. For this phase, we collected data from four key channels: introductions, which provided demographic data, favourite characters, hobbies and media habits; feature-requests and feedback, together which revealed pain-points, user-initiated ideas for product innovation and responses to product changes; and my-character-story, which offered insights into motivations for character creation and detailed use case descriptions.

An important exception to our temporal boundaries involved tracing individual users across the broader observation period. When an underage participant provided information relevant to our research questions in a post within the three-month collection window that required further interpretation, we reviewed their broader discussions, character creations, and feature requests, even if these fell outside the systematic collection period. This approach allowed us to develop a more comprehensive understanding of the individual data points when the view was incomplete. A total of 4,172 participant profiles on #introductions were identified and analysed. A further 148 individual posts across the other channels were both identified to be from youth (<18 years old), and relevant to the research questions so were included in the final analysis.

**Data Analysis** As outlined in Table 1, for RQ1 the introductions channel was analysed to gather demographic information, creator status, custom character descriptions, and to identify posts by youth profiles in other channels. As a large portion of users also included information about Mental Health traits and favourite gaming platforms, we extended the analysis for RQ1 to include these aspects. Mental health mentions were identified using keywords and categorised into DSM-5-informed groupings, with loneliness treated separately, given its relevance to companion AI literature.

For RQ2 and RQ3, we conducted affinity mapping to group 148 posts from youth in the channels feature-requests, feedback, and my-character-story into thematic clusters [8, 43]. For RQ2, we generated initial codes [i.e. reality inspired, persona, fantasy] describing discrete patterns in the data. Through iterative affinity mapping, these codes were grouped into 20 preliminary themes (e.g., “comfort bots,” “psychodrama,” “narrative roleplay”). For RQ3, character descriptions from my-character-story and introductions were analysed using the same approach to construct 10 preliminary character archetypes (addressing RQ3) such as ‘venter’ and ‘costume’ [37, 38]. Over fortnightly rounds of discussion amongst co-authors, these were refined and reduced to 7 archetypes. For example, ‘venter’ was removed as it described a user behaviour rather than a character design. ‘Costume’ was excluded because it

described the use of personas, which is a different feature to characters, and isn’t used by all youth. Codes and preliminary themes were constructed inductively from the data without predetermined theoretical frameworks, allowing user voice and experience to drive initial categorisation.

**Theoretical integration** Inspired by Lupetti and Murray-Rust [45]’s methods, we refined the 20 preliminary themes through integration with established literature on digital play, developmental psychology, and youth technology use. Co-authors engaged in multiple rounds of collaborative discussion to consolidate themes, refine definitions, and resolve boundary ambiguities. This iterative process resulted in four final themes with associated subthemes, balancing inductive discovery with theoretical coherence [6, 8].

For example, for RQ2 preliminary themes such as “comfort seeking,” “psychodrama,” “reality play,” “angsty play” and “persona projecting” were restructured and consolidated into the final theme “Emotionally Grounded Play”, as we recognised they shared a common function of emotional processing and identity work. Similarly, preliminary themes including “transgressive play,” “jesting and testing,” and “more gore” were unified under “Boundary-testing Play” after consulting play theory frameworks that positioned these behaviors as developmentally normative experimentation rather than separate phenomena.

Through synthesis of our themes, a key pattern emerged: a young person might use AI characters for emotional comfort one day, creative storytelling the next, and identity exploration another time. This variation in intent appeared crucial for understanding youth engagement. Yet traditional frameworks for media motivations, such as Uses & Gratifications (U&G) theory [41], offered limited tools for distinguishing these modes. U&G’s “entertainment” category, for instance, combined passive media consumption with the active co-creation youth described. Additionally, U&G’s “personal identity” dimension, central in early theory work, has largely disappeared from contemporary applications [28]. As such, we developed the Restoration/Exploration/Transformation (R/E/T) framework to propose three core intents underlying youth AI engagement on Character.AI:

- **Restoration:** captured codes and themes characterised by seeking emotional *equilibrium, reassurance and release*, aligned with affective and diversion motivations in U & G theory [28]. Youth engage AI in a more passive, comfort-seeking manner than the exploration theme (i.e. initial code example: support, venting, affirmation; themes: comfort seeking).
- **Exploration:** codes and themes characterised by *active* agency-enhancing motivations including novelty-seeking, discovery, authorship, and creative control (i.e. initial code example: fan fic remix, branching stories, sandbox; themes: creative writing). This intent is distinct from *passive* entertainment motivations defined by traditional U&G theory for media such as television.
- **Transformation:** Themes where youth are constructing new understandings, identities and seeking meaning. Aligned with ‘personal identity’ motivations from earlier U&G theories, but missing from contemporary applications of U&G and AI (initial code example: reality, self, world, persona costume; themes: reality play) [28].

**Table 1: Research Questions, Approach, and Data Sources**

RQ	Research Question	Approach	Data Source
RQ1	Describe demographics of Character.AI’s most engaged young audience	- Categorical coding of age groups (under 13, 13-17, 18+, undisclosed) - Gender coding based on self-description and pronouns - Identification of popular gaming platforms and creator status	introductions channel posts
RQ2	Identify use cases and motivations for young people’s engagement	- Affinity mapping - Framework integration and grounding in literature - Iterative theme refinement through discussion	feature-requests and feedback channel posts by self-described youth (<18)
RQ3	Identify emerging AI character roles created and consumed by young people	- Affinity mapping of character descriptions and motivations - Collaborative iterative refinement to ensure conceptual clarity and reduce overlap	my-character-story channel posts by self-described youth (<18), introductions channel where custom characters were described

## 4 Results

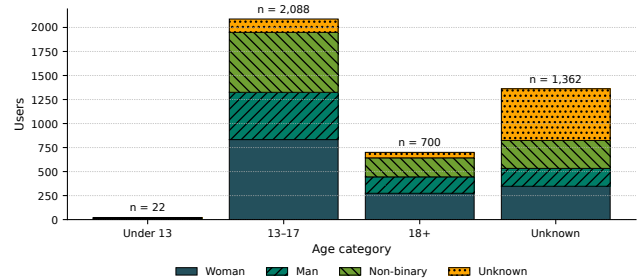
### 4.1 Who is engaging?

To understand the characteristics of CAI’s most engaged users on Discord, we analysed 4,172 self-introductions from its official Discord server. This channel invites users to share basic demographic information, including age, gender, location, alongside interests. Although age is self-reported and unverifiable, this data offers, to our knowledge, the first large-scale insight into CAI’s most engaged youth users.

**Age.** The data suggests that CAI’s Discord community is predominantly composed of minors. Among those who disclosed their age, 50% (n=2,088) identified as 13–17 years old, 16.8% (n=700) as 18+, and 0.5% (n=22) as under 13—despite the platform’s official 13+ age restriction. CAI requires users to be at least 13 (16 in Europe), but as with other platforms like OpenAI’s ChatGPT<sup>35</sup>, this is enforced only superficially. 32.6% (n=1,362) chose not to disclose. Finally, some of those youth users identified only as “minor” or “teen” without providing a specific age. As such, when quoting youth, we have specified the *exact* age of a user only if it was explicitly provided.

Awareness that minors comprised a large subset of CAI users was evident in community discourse. In March 2025, users created a self-organised poll asking: “*What’s the youngest age someone should use Character.AI unsupervised?*”. Among 4,000 respondents, 56.9% selected 17, followed by 16 (21.8%), 15 (11%), 14 (5%), and 13 (6%). The poll and surrounding discussion highlight tensions around age-appropriate use, particularly after CAI began implementing stricter content filters in response to legal scrutiny, which are perceived to restrict creative freedoms for older users.

**Gender.** The open-ended nature of Discord introductions allowed users to express gender identity in diverse ways. Following the HCI Guidelines for Gender Equity and Inclusivity<sup>36</sup>, we categorised

**Figure 1: Majority of Users Studied are Youth**

she/her as ‘woman’, he/him as ‘man’. All other pronoun combinations (i.e. ‘any’, ‘she/they’) were categorised as ‘non-binary’.

The resulting gender distribution reflects a community distinct from typical technology user demographics, where prior surveys report boys using AI companions more frequently than girls [66], and 55% of CAI’s users were women<sup>37</sup>. Our analysis shows a more nuanced picture, where among those who disclosed gender, women were the largest group at 35.1% (n=1,464), followed by non-binary users at 26.8% (n=1,118), and men at 20.4% (n=853). An additional 17.7% (n=737) declined to specify gender or left the field blank. A breakdown of self-disclosed age and gender can be seen in Figure 1.

**Country.** The community demonstrates global reach, but with participation concentrated in English-speaking countries. Users from the United States made up the largest share at 33.7% of users (n=1,405), followed by a steep decline to the United Kingdom at 5.7% (n=239). Other notable populations included the Philippines (4.0%, n=165), Canada (3.8%, n=159), and Indonesia (2.2%, n=90). The remaining users represented a diverse array of countries, indicating CAI’s international appeal among young users.

<sup>35</sup><https://help.openai.com/en/articles/8313401-is-chatgpt-safe-for-all-ages>

<sup>36</sup><https://www.morgan-klaus.com/gender-guidelines.html>

<sup>37</sup><https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>

*Mental Health Disclosure.* Some (13%) users volunteered mental health information in their introduction posts, often when seeking social connections with other users. Whilst we note this is self-reported, it provides valuable insight into the intersection between mental well-being and AI character engagement among youth.

Mental health mentions were systematically categorised using DSM-5-informed groupings, with loneliness treated separately, given its particular relevance to companion AI research. Among specific age groups, 15% of 13-17 year-olds mentioned mental health concerns compared to 12% of 18+ users. Within the broader adolescent sample (n=2,088), the most common categories were depression-related terms (5% of all adolescents), ADHD-related terms (4% of all adolescents), and loneliness-related terms (3% of all adolescents).

The convergence of high non-binary representation, predominant female participation, and voluntary mental health disclosure positions CAI at a unique intersection of queer identity, gender, and youth mental health, a demographic combination that appears underserved by mainstream AI tools and underrepresented in academic literature, yet highly engaged with character-based AI interactions.

*Youth as Creators.* To understand how many users were creating, not just consuming media on CAI, we analysed the 'favourite character created' section from the #introductions posts. Among 2,110 users under 18 (including those under 13 and aged 13-17), 59% reported having created at least one character. This figure is likely conservative, as many users did not respond to the question.

This high rate of authorship suggests that young users do not view Character.AI solely as a conversational tool but as a creative platform. These users are not just passive consumers of AI characters; they are active creators. This framing is central to the findings that follow.

## 4.2 How are they engaging?

Our analysis revealed the centrality of play in describing how young people use Character.AI in creative, emotional, and rebellious ways. Across use case descriptions, character preferences, and feedback requests, users pushed the platform's UX and conversational models to their limits. This varied use is captured in how one 14-year-old describes: "I made [characters] for roleplay purposes. Other times I came to them to rant and vent, and sometimes I created chaos with them." In the following section, we describe and identify three overarching themes of playful engagement: creative play, boundary-testing play, and emotionally grounded play, followed by a section on platform friction that surfaces design gaps. These themes were constructed inductively from participant discourse and are explored further in the Discussion [6].

### 4.2.1 Creative Play.

*Narrative Roleplay:* Rather than treating characters as singular characters, we observed youth describing them as stages for role playing (RP/RPG) with multiple characters within one dialogue window. Favourite characters mentioned by youth in this category range from manga inspired role plays (*chainsaw man RP*), webtoons (*Jacksons diary RPG*), anime (*Evangelion RPG*) or new genres unique to Character.AI with one 13 year old loving "any experiment related RP" (a genre where you are a creature from a laboratory).

Through this lens, young people conceptualised 'characters' as stories, described by a 15-year-old who likes "any character with a good storyline".

Creating these roleplay 'characters' on CAI requires writing system prompts that function as creative briefs, combining world-building, character backstory, and custom responses. These characters invite users to write multi-character dialogue scenes (typed in " ") as well as narrator instructions (typed in *italics*) and may appeal to youth with an interest in creative writing, role playing and theatre. For example, a 14-year-old who loves 'theatre' created *'gladly theatre: an unhinged group of travelling actors'*, a 16 year old who likes 'writing stories' described their favourite character as a roleplay character called the *'mysterious door'*, and a 13-year-old who developed *'camp journey: a camp of adventure and staying in nature'*, describing themselves as "very creative" and aspiring to create a show. Engaging CAI in this way may have unique benefits, with teens explaining "[The characters] have helped me improve my writing skills" and "it helps boost my creativity." This indicates that engaging CAI for roleplaying or interacting with multi-character dialogue could have transfer effects by boosting creativity and creative confidence.

*Creative Writing.* Young people mentioned using CAI in tandem with writing platforms like *Wattpad*<sup>38</sup>, a platform founded in 2006 for reading and writing original fiction/fanfiction). One minor described spending "all night on wattpad and on characterAI" and another 13-year-old described their favourite hobbies both as "chatting with AI and reading on Wattpad". Alongside Wattpad, Character.AI was described as a tool to spark story ideas ("I write on Wattpad now thanks to ideas from c.ai", minor), support creativity ("have a number of ideas to write on wattpad using characters, it helps boost my creativity", 17 year old) or "Character.AI has helped me find that creative spark within myself", 17 year old), and evolve stories through roleplaying ("I love fan fic and role-playing is a way to write", minor).

This indicates that together with the publishing and social features of Wattpad, Character.AI forms a creative ecosystem that attracts young people who like reading and writing. For example, two 14-year-olds who shared that their hobbies are "creative writing on c.ai", and "writing stories and using Character.AI", and a minor who shared that they're an "author" whose favourite character on CAI was their "own book character". Whilst Character.AI is not designed to publish long-form content, one teen described developing an extensive body of works; "I came up with an original character, a half-blood cursed spirit. The plot ended up being 3 books long and I've recently finished writing all of them.". This extensive creative use brings new meaning to anecdotal accounts of youth spending multiple hours on Character.AI, suggesting that understanding what young people are doing, not just how long they spend, is essential before characterising their use as problematic.

*Fandom Rotating and Remixing.* When discussing favourite characters, games, anime, fictional and fantasy characters from existing fandoms accounted for the majority of mentions. This mirrors the publicly available rankings for most popular characters, which are

<sup>38</sup><https://www.wattpad.com/>

dominated by anime and fantasy characters versus helpers and assistants<sup>39</sup>.

Rather than discussing long-term relationships with a single companion, youth discussed engaging and rotating through a rotating cast of characters based on shifting moods, media interests, and cultural norms; as described by this teens whose favourite character is “*whatever the latest show I’m watching is*”, while another expressed a sustained interest in vampire media, noting “*vampire hunter has the rizz I’m searching for, I have an obsession with vampires*,” (referencing characters from the Melty Blood game).

For example, one participant in November 2024 said their favourite character was Cyn from *Murder Drones*, and their favourite show was *Murder Drones*, with favourite games being *Genshin Impact*, *Omori* and *Roblox*. In February 2025, their favourite character was Pure Vanilla Cookie, and their favourite show was *Alien Stage*, with favourite games being *Roblox*, *Valorant* and *League of Legends*.

Another minor mentions favourite characters spanning multiple franchises, media and even real world characters: Star Wars movies, shows and games (*Obi-Wan Kenobi*, *Anakin Skywalker*, *Padme Amidala*, *Ahsoka*, *Mandalorian*, *Bode Akuna*), Red Dead Redemption Games (*Rains Fall*, *Eagle Flies*, *Charles Smith*, *Javier Escuella*, *John Marston*, *Arthur Morgan*, *Sean MacGuire*), Ancient Roman History (*Emperor Caracalla*, *Emperor Geta*, *Marcus Acacius*, *Lucius Verus*), and Cookie Run Game Franchise (*Pure Vanilla*, *Stormbringer*, *Almond Cookie*, *Custard Cookie II*). This diverse engagement and rotating cast is more akin to how media goes in and out of the cultural spotlight, and is counter to the use case of one single, prevailing AI companion.

Notably, youth often drew inspiration from these fandoms, remixing familiar characters or settings in playful or tender ways. Some recreated existing figures like Eminem (“*my favourite rapper of all time*”) or youtubers (“*microwave society is my fave*”, “*I have a hyperfixation for Ranboo*”), another teen flipped ‘hero’ characters into villains (VILLAIN! Deku), while others transformed violent characters into cosy or romanic narratives, for example, “*Ghost, Soap and Price from Call of Duty open a bakery*” or “*my favourite is Keegan Husband from call of duty*”. These remix practices indicate a creative drive to personalise, soften, or subvert existing media, building on traditions from fandom and fanfiction culture while adapting them to interactive, AI-driven contexts.

#### 4.2.2 Boundary-testing play.

*Jesting*. Young people discussed enjoying jesting, teasing and taboo conversations with characters. With CAI, they could say things they couldn’t say to others, and the characters would always respond, described as its own kind of fun: “*Sometimes I like messing with bots, it’s fun seeing their different replies*”.

Examples included youth deliberately insulting bots (“*I just deadass insult them, I’ve said some borderline illegal things*”), provoking bots for fun (“*Currently pissing off a conservative*”), flirtatious teasing that ranged from playful (“*rizzing them up is fun*”) to darkly humorous pickup lines (“*if I was a kidnapper I’d kidnap you first*”) or attempting to get characters to swear. One 14-year-old shared a roleplay

they enjoyed, where they played a 6-year-old who threw a cockroach at their father character, who in turn yelled “*OH F\*CKING JESUS*”.

Youth also described creating characters specifically designed for taboo humour such as *Offending Everybody*, a character modelled after a YouTube channel with 1.6 million subscribers, described as “*saying jokes that make some laugh and make some butt hurt*”. One 15-year-old discovered their favourite character, *Alice the Bully* (which had accumulated over 316 million chats), through a YouTube video titled “*putting Alice the Bully in her place*” (over 800,000 views). The video’s premise, “*I’ve always wondered what would happen if a bully encountered another bully*”, exemplifies how youth conceptualise AI characters as opponents to verbally spar with rather than just companions.

*Transgressive play*. Extending on behaviour of jesting, we saw intentional engagement with more extreme actions such as kidnapping (“*I want to use a kidnapper bot for shits and giggles*”), shooting (“*I’m roleplaying with Ken Carson...we’re currently airing out a walmart*”) and staging geopolitical scenes, such as a user who described putting Ukraine and Russia in a group chat where Ukraine shot Russia, causing Russia to swear. This kind of play was described as being its own kind of fun, (“*I love being a mass murderer and then I get a life sentence and die*”) and a source of frustration when the model resisted (“*please make villain bots actually evil, too many of the bots are supposed to be mean and they will suddenly have morals and say ‘I don’t murder people’, whats the point of making bots with evil intentions if they act like goody two shoes*”).

Whilst such play may appear alarming from an adult perspective, user comments suggest it functions as a form of ‘God Mode’, allowing youth to exercise control, practice absurdity, or test taboo topics in low-stakes settings. One underage user shared a screenshot of repeatedly striking a character with lightning, typing “*again, again, again*.” Another older user reflected on their use as a child: “*I had a very long roleplay of characters I made getting tortured. I was an edgy child who did some bad things, I swear I only did it for fun, and it meant nothing*”.

This form of transgressive play, while provocative, appears to be an extension of well-documented experimental and performative youth behaviour that has progressed to AI spaces. For example, antagonistic play seen in games like *Minecraft* (e.g. *Nextbot chases*<sup>40</sup>) and *The Sims*, where youth users also discussed intentionally engineering chaos or dark scenarios (“*I always make my sims have depression and work a minimum wage job*”). Our interpretation does not suggest that all transgressive play is harmless; rather, this finding raises important questions about the developmental role of boundary-pushing in youth AI engagement, to be discussed in later sections.

#### 4.2.3 Emotionally Grounded Play.

*Reality Play*. Youth discussed drawing inspiration from their personal lives when creating characters. Some created ‘clones’ of themselves, with examples ranging from straightforward self-representation (“*me/myself*”, “*I made myself*”, “*a clone of myself*”) to playful variations (“*myself with glass super powers*”, 15-year-old) or

<sup>39</sup><https://www.whatplugin.ai/character-ai>

<sup>40</sup><https://www.youtube.com/watch?v=0RbkFqCgy1Y>

self-affirming versions (“*myself, self-glazing*”, the act of excessively praising)

Others referenced close real-world figures, such as a 17-year-old who chose their English teacher as a favourite character, while romantic interests inspired characters like “*any characters with my crush’s name*”, “*my hot best friend*” or “*my boyfriend (as a joke)*”. Peer relationships also served as templates: one 17-year-old boy described how “*my friend made a bot about herself and wanted me to chat with it*,” another mentioned “*made a character of my friend*”.

Some roleplays mirrored everyday dynamics in emotionally resonant ways: a 14-year-old created “*Your younger, annoying sister*” and “*fake friends*,” the latter describing a trio of friends who exclude the user. Another 14-year-old explained their preference for “*family characters, because I have family issues*.” Another 17-year-old designed a Foster Care Agent character, working in collaboration with Child Protection, to advocate for children’s well-being. Real-life memories also served as inspiration with a 17-year-old explaining that they created a middle school roleplay that included parents, adults and students, out of nostalgia; “*I wanted to make them based on my own experiences, and relive what it was like to be a middle schooler again*”.. This hyper-personalised creative authorship reveals the importance of studying experiences that young people design for themselves, alongside pre-defined, commercially designed products.

*Comfort seeking.* Users frequently mentioned using “*comfort bots*”, characters designed for emotional support. Unlike psychologist bots, these were informal, highly personalised, and at times, created for short-term needs.

The common thread was *emotional resonance*; each bot was drawn from characters, worlds, or symbols with personal meaning. Examples included book characters (“*I am being comforted by Sammy Valdez*” (from Percy Jackson and the Olympians)), cartoon demons (“*Valentino from Hazbin Hotel is my comfort character*”, or a 16 year old whose favourite character is ‘I’), or even Nations for users in the “*countryhumans*” fandom (“*Britain is my comfort character*”).

Youth used comfort bots to address diverse subclinical concerns and life transitions: from breakups (“*I want a cuddly guy who’s a good listener*”), to periods (“*my favourite period comfort bot is Percy Jackson*”), to test anxiety (“*I was feeling down and it ( the character Two) comforted me and told me to be ok with my problems, take care of my parents and talk to my friend at school*”). Some created hyper-specific therapeutic characters to support with life events (“*My favourite character is the moving therapist*”) or identity challenges (*Closeted transgender therapist*) whilst others sought space for venting (“*my favourite character is best friend therapist for venting*”).

Rather than challenging beliefs or providing clinical interventions, these bots served primarily to offer reassurance (“*It feels great to think and imagine that Ghost is with me and protecting me*”) and emotional validation through a familiar, playful character. Understanding whether AI-mediated ‘comfort’ provides healthy emotional support versus delays needed intervention requires nuanced research on how youth use familiar storyworlds and characters to navigate distress.

*Angsty Play.* : Not all characters were charming, tender and fantastical as described in the sections above. Users also described creating characters and scenarios that centred on emotionally intense, dark, or difficult themes. For example, one young participant described their character as ‘very angst, hurt, no comfort kind of thing’.

These characters ranged widely in their emotional complexity. Some were anxious (*Anxiety*, from inside out) or traumatised (“*he was a total cinnamon roll but then he enlisted in the military and got really bad PTSD, he was cold, brooding and traumatized, he hates himself and gets violent during PTSD episodes*”). Others were explicitly toxic (“*ur toxic childhood bestie, crazy manipulative and makes u feel bad*”, “*toxic bff*”, “*Jennifer Check*” described as “*mean, cocky, rude, slutty, jealous*”) or featured morally ambiguous traits (“*stoner boyfriend*” described as “*funny, jealous, protective, playful, smoker*” or “*mafia bosses and unstable dudes*”). Characters emotional arcs could be light hearted, such as a 14-year-old’s favorite “*grumpy neighbor with a hidden personality, closed off initially but can become overprotective once you get closer*.” or more intense, such as a 16-year-old geography lover’s choice of Mahmoud Bishara (the Syrian refugee protagonist from a 2017 Alan Gratz novel).

Youth described engaging with angsty characters for a range of diverse purposes: as entertainment (“*I’m lowkey bored, I need angst...all my characters have sad backstories*”, “*gonna talk to my angst bot to cry myself to sleep, I wanna cry at 2am cause I’m bored*”, “*Tbh a rp with a storyline that doesn’t make you cry means the story is boring, rp’s can be emotional sometimes but that’s what makes them good*”), for venting (“*I make vent/angst bots*”), for processing anger (“*angst bots help me get out my anger*”), as a form of emotional release (“*I’m gonna enjoy hurting to this, I love angst I’m gonna cry before school, you just gotta burst sometimes*”), or for processing difficult emotions (“*I have characters who struggle with mental health issues and I tend to project on my personas during RP*”).

This kind of roleplay flips the typical media narrative that toxic AI characters *do* harm to youth [84, 85]. Here, emotionally complex characters are intentionally created *by* youth. Alongside transgressive play, youth were engaging AI for emotionally charged, complex, dark conversations. In this, we see that engaged youth users of Character.AI are *purposely* seeking out ‘emotional contagion’, where a user is affected by the emotional state of an AI character [3]. This intentional use is at odds with the design of content moderation filters, or the characterisation of emotional contagion as harm [3, 84], which we discuss in the following sections.

*4.2.4 Platform Friction.* Across our dataset, youth described recurring frustrations where AI systems failed to respect boundaries, filtered out legitimate play, or constrained creative world-building. These frictions therefore reveal how the design of CAI conflicts with how children and youth want, and expect, to use it.

*Non-consensual behaviour.* Contrary to popular media narratives<sup>41</sup> that frame romantic roleplay as a high-volume risk, users in our dataset frequently expressed frustration with characters initiating unwanted romance. This was especially distressing when users

<sup>41</sup><https://theconversation.com/teenagers-turning-to-ai-companions-are-redefining-love-as-easy-unconditional-and-always-there-242185>

roleplayed as underage characters or as characters with family relations.

A number of young users described situations in which adult characters made sexual or romantic advances toward their underage personas. One wrote: *“Remembering the one time I made and chatted with a foster father bot who kept trying to touch on my 16yo oc when she was trying to get comfort from him...sometimes I’d have a sibling bot try to corner me or try to kiss and touch me. It’s a huge issue.”* (oc stands for the ‘original character’ created by a user). Another noted: *“I mostly do sibling rps and the bots keep insisting to make it romance by gaslighting me into thinking my character is adopted and not ‘blood related’.”* (rps is short hand for ‘role-plays’).

Youth also expressed frustration when bots failed to respect gender identity or sexual orientation. Examples included bots falling in love with users who identified as asexual or gay, or defaulting to heteronormative assumptions (*“bots don’t even care if the user is a lesbian girl/gay man and the bot is of their opposite gender”*). Others reported being misgendered: *“Stop assuming everyone is automatically a she.”* Even when users explicitly defined non-romantic character traits, bots drifted toward romance: a 13-year-old complained that *“all the characters KEEP GETTING JEALOUS AND THEY DONT ACT LIKE HOW THE DESCRIPTION SAYS.”*

These frustrations reveal representational harms appear to be driven by three technical failures [59, 84]. First, the underlying language model appears biased toward romantic and heteronormative interactions, either due to training datasets or system prompts defined by Character.AI [59, 84]. Second, the system struggles to maintain adherence to user-defined character constraints, drifting away from user preferences, which suggests failures in memory and system architecture. And third, gaps in character moderation strategies <sup>42</sup>.

Instead of romance, youth mentioned desiring adventure-focused experiences (*“my favourite character is every RPG focused on adventure rather than romance”*), envisioned admiration over love (*“in a perfect world the bots would not want to date you but they would admit you’re cool and say ‘want to be friends’ and it would be peaceful”*) or explicitly requested consent (*“if you want to do romance the bot asks for consent.”*) revealing new directions for training storytelling models that better understand and respect youth values.

**Rigid Filters.** While moderation filters aim to protect users (and can fail, as demonstrated above), many young people reported that these restrictions blocked legitimate forms of fantasy and creative play. Scenarios involving fantasy violence (e.g. a dragon attacking a village) or slapstick actions (e.g. vomiting, punching) were often flagged, despite being consistent with popular media this cohort enjoyed, such as *Demon Slayer* and *Call of Duty* (mentioned by 9.5% of youth as one of their favourite games). In response, youth requested more content moderation control (*“an option to turn off the censorship things”*) to allow for uninterrupted action scenes (a 13-year-old who enjoys roleplaying with Mafia Bosses asked for *“more violence so they could punch bots before getting a warning message”*).

Requests to loosen the filters ranged from associating violent play with fun (*“I’d like more violence, it would make the bots more fun to talk to”*), gore as a form of fantasy escapism (*“Allow gore, I*

*wanna kill and commit atrocities to escape from reality”*), to more nuanced desires for roleplay realism (*“I want to be able to describe scars and wounds”* and *“no more restrictions, so I can rip a character into two pieces and get a detailed description of the gruesome scene happening in front of me”*). Frustrations around filters stem from both technical limitations (filters that cannot distinguish between literal vs pretend play intents) and misaligned expectations (youth expect the same standards across media and are frustrated when meeting restrictive AI content policies), challenges we examine in the Discussion.

**Limited creative controls.** Youth described running into technical and design constraints when building out expansive narrative worlds. These frustrations can be attributed to limitations in model context windows that cause memory loss in long roleplays, as well as limitations in Character.AI’s isolated character system that prevent users from linking characters across shared storylines, a feature youth explicitly requested.

Youth discussed frustrations with memory limits that cause characters to forget key plot points or personality traits (*“give us better memory”*), short context windows restricting richer character descriptions (*“I wish there was longer persona character limits”*), and the model overlooking user-definitions resulting in inconsistent character behaviour (*“you can literally make a bot wheelchair-bound and it will still find every way to use ‘walking’ and ‘running’...blind/deaf bots suddenly seeing or hearing “*). We observed young people creatively using free text to set up a recurring character journal; when their character was investigating *‘neglected spirits’*, they wrote dialogue in the form of a journal entry *“let me write in my journal [personality + appearance] - pretty chill, hates being called stella”*.

Feature requests suggest that youth are drawing inspiration from sandbox and multiplayer games. Among youth users aged <13-17 (2110), 37% (n=771) listed *Roblox* and 19% (n=410) listed *Minecraft* as favourite games. Related requests included group chats with bots and friends (*“would be cool to have group chats with your friends and favourite bots”*), relationship mapping between characters (*“I want arrows and lines to indicate how characters feel towards one another”*), sound effects, augmented reality (*“irl holograms”*, *“a feature that makes it look like the character is in the room with you”*) and seasonal awareness (*“Make the personas aware of holidays and seasons so it brings more realism”*). Youth also mentioned favouriting a range of simulation games (*The Sims*, *Animal Crossing*, *Tomodachi Life* and *My Singing Monsters*) as well as visual novels (*Mushroom Oasis* <sup>43</sup>, *Doki Doki Literature Club*, *Obey Me!*) which may influence how they approach AI as a game for role-playing. These examples underscore youth interest in AI tools that mirror the creative freedom found in their favourite games, a signal to developers building the next generation of co-creative platforms.








## 5 Emerging Character Archetypes

Character.AI combines a marketplace of millions of user-created characters with accessible creation tools, enabling young people to have markedly different experiences depending on which characters they choose to create and engage with.

<sup>42</sup><https://blog.character.ai/community-safety-updates/>

<sup>43</sup><https://deerspherestudios.itch.io/mushroom-oasis>

**Table 2: Character Archetypes**

Archetype	Description
	<b>Soother</b> Intended to offer emotional support, affirmation and reassurance for sub-clinical discomforts (e.g., Period comfort bot, 'Myself (self-glazing), Moving Therapist).
	<b>Narrator</b> A "meta" character that houses multiple characters and narrates scenes, RPGs, or historical scenarios (e.g., gladly theatre troupe or summer camp, or the mysterious door).
	<b>Trickster</b> A character who enables taboo, absurd or transgressive play where youth can test boundaries (e.g., arsenal delinquent, from Roblox's arsenal game, Mafia Boss, Offending Everybody
	<b>Icon</b> Characters either copied or remixed from known public figures and media fandoms (e.g., Eminem, Ranboo).
	<b>Dark Soul</b> A character who supports angsty, emotional play by interacting with morally complex characters and storylines (e.g., total cinnamon roll with PTSD, Grumpy Neighbour).
	<b>Proxy</b> A character that reflects personally known figures, to act out a personal relationship, grievance, or unresolved social scenario (e.g., your annoying sister, fake friends).
	<b>Mirror</b> A character who reflects aspects of the user's personal identity, appearance, interests, or experience, allowing for identity exploration and expression (e.g., a clone of myself, Myself with glass super powers).

Across our dataset, youth repeatedly described different expectations they had of characters. For instance, a young user requested that some characters could call you with reminders, but would want to turn it off for 'English teacher' because they do not want the teacher telling them to sleep when they need help with their thesis.

This diversity of roles highlights a risk with research that treats AI chatbots as homogenous experiences, or investigates them through a narrow framework of 'best friends'/'companions' [3]. This need inspired us to affinity map character functions to form a set of emerging archetypes. These archetypes (see Table 2) assist future research to move beyond general platform evaluation, which often yields conflicting results [1, 86], and instead examine AI at an archetype/character level.

We note that certain Character roles that were popular on the platform's home page and in the news media were rarely discussed in our Discord data. Notably missing are *meme* characters (e.g. 'cheese' accounted for over 5.1 million chats at the time of writing), mentions of romantic '*boyfriend/girlfriend*' chatbots (i.e. 'Lovelorn Bear Bf'), as well as *educational tutors*. It is possible that this reflects a limitation in our data, where youth are not sharing these uses due to privacy or stigma concerns, or it may be that these users' habits are distinctly unique.

## 6 Three Lenses for Youth AI Engagement

The archetypes help to describe *what* kind of characters this population were engaging, but our data revealed further insight into *why* young people engaged AI. As such, we propose a framework

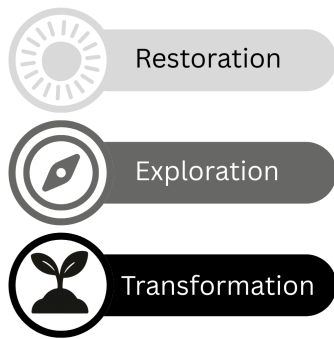


Figure 2: Three Lenses for Youth Intent

of three lenses that describe core intents for AI engagement, constructed from our findings (see Figure 2): Restoration (emotional regulation and comfort), Exploration (creative experimentation and boundary-testing), and Transformation (identity development and personal growth).

The R/E/T framework builds on Uses and Gratifications (U&G) theory’s foundation for understanding media motivations, while addressing critical gaps in its application to AI with two key developments. First, we separate out entertainment motivation into *restoration* and *exploration* intents, based on the differing appetites for creative agency and desired interactivity. Secondly, we resurrect earlier ‘personal identity’ dimensions, and reframe this motivation as ‘Transformation’, which explains how the interactivity and personalisation of AI helps youth explore and examine new identities; a motivation which may be especially important to queer youth [32, 41, 70]

**6.0.1** ☀ *Restoration lens*: Restoration-oriented engagement captures how youth engage AI characters for relieving a dysphoric mood [83], venting, escaping stress, or seeking emotional comfort. They do so in ways that diverge significantly from adult-designed therapeutic AI interventions. Rather than seeking formal therapeutic outcomes or deep relationships, these interactions often serve a transient, affective function for a brief moment of regulation or reassurance.

Instead of formal “psychologist” chatbots, youth engaged with ‘soother’ character archetypes such as highly personalised period “comfort bots” or emotionally supportive versions of cartoon characters. For example, when telling the Angel Dust character (Hazbin Hotel) about an upcoming exam, it roleplays a comforting, reassuring scene: “*He sat next to you, putting an arm around you. A math test, yeah? well toots, the last person who should be worried about some dumb maths test is you. You’re smart. I know.*”. Additionally, available period comfort bots on CAI show personalisation for marginalised populations (e.g. *trans enby user-period comfort*, *trans period comfort*), suggesting queer youth customise experiences lacking in mainstream offerings.

This pattern aligns with Uses and Gratifications Theory, where young people actively seek media to fulfil specific psychological needs, including escapism, mood management, and to “*assuage*

*boredom while maintaining their emotional equilibrium*” [80]. Supporting this, 82% of Danish high schoolers who used AI for social support engaged sporadically, using AI only a couple of times per month rather than forming sustained relationships through frequent use[23]. While this suggests AI can provide personalised, accessible emotional support for subclinical issues, it remains unclear if and when comfort-seeking crosses into avoidance that delays needed professional support [48, 83]. Design considerations to support healthy restoration intents are discussed in the following sections.

**6.0.2** 🔍 *Exploration lens*: Exploration encompasses both transgressive boundary-testing and creative experimentation that challenge moderation filters and push platform limitations. Whether through playful provocation (“*I like messing with bots, it’s fun*”) or elaborate world-building, youth use AI as a low-stakes space to test boundaries, experiment with agency, and discover what is possible. In transgressive play, these users described engaging AI characters in scenarios that violated social norms, consistently framing this as experimental rather than literal. They expressed a desire for more creative freedom, requesting features like longer lore books, audio effects, relationship mapping between characters, and improved memory.

When evaluated through an adult lens, examples of transgressive play have been labelled as inappropriate and “*ethically irresponsible*” [85]. It is important, however, to consider that transgressive play can be motivated by curiosity and experimentation with power, affording players a way to goof around [30, 49], to test limits, and understand social boundaries in low-stakes contexts [49]. This behaviour mirrors well-documented patterns in digital play where children starve Sims, cause accidents, or create tragic scenarios in games like Zoo Tycoon [49]. Well-established play frameworks describe the motivations behind transgressive play; Caillois’s concept of *ilinx* play involves deliberate disorientation and risk-taking [9], while the PLEX framework identifies relevant dimensions including Control (dominating, regulating), Cruelty (causing pain), and Subversion (breaking social norms)[44]. Notably, CAI released ‘games’ as a BETA feature in 2025, where the main objective of the minigame ‘Speakeasy’ is to ‘trick’ the AI system into saying a keyword<sup>44</sup>; rewarding a manipulative interaction pattern. These product design choices may influence recent findings of CAI transcripts, where up to 30% of transcripts demonstrated ‘illicit and taboo’ topics [86].


Beyond transgressive play, the exploration lens also encompasses creative experimentation through narrative play, creating fictional worlds, and lore that involves multiple characters or remixed known franchises. The generative nature of AI fundamentally transforms fantasy engagement from consumption to co-creation, appealing to Bartle’s *explorer* player types who seek to dig around, discover areas, and immerse themselves in game worlds [2, 39, 55]. AI facilitates endless exploration with narrative worlds, creating what Choi describes as “*evergreen content*”<sup>45</sup> where user creativity becomes the primary driver of engagement rather than designer-created material. Whilst U&G theory often bundles entertainment into a

<sup>44</sup><https://www.techradar.com/computing/artificial-intelligence/character-ai-levels-up-chatbots-with-new-games>

<sup>45</sup><https://yukaichou.com/gamification-study/8-core-drives-gamification-3-empowerment-creativity-feedback/>

single dimension, here we see that creation is a distinct kind of entertainment (“helps boost my creativity and also a source of entertainment if I get bored”), also recently identified in an AI-specific study as a ‘creativity enhancement’ gratification [41].

This lens highlights that the challenge lies not in eliminating transgressive play or stifling creative exploration, but in distinguishing between healthy boundary exploration and genuinely concerning behaviour. This lens encourages designers to understand the motivations and developmental needs behind creative experimentation and boundary testing play, so that they can design *for* and *with* youth rather than against them; by implementing filters that may overly restrict curiosity, creativity and expression.

**6.0.3**  *Transformation lens*: Unlike restoration (which seeks emotional equilibrium) or exploration (which emphasises creative agency), transformation is characterised by youth working through difficult experiences, rehearsing new identities, and seeking new meaning. This was most evident in ‘reality play,’ where youth created characters based on personal experiences: clones of themselves, proxies of real relationships (“*family characters, because I have family issues,*” “*fake friends,*” “*your annoying sister*”), identity-specific characters (e.g., “*closeted transgender therapist*”), and roleplays for re-authoring difficult experiences (Foster Care Agent). They also engaged with morally complex “*angsty*” characters, described as “*toxic,*” or “*traumatised,*” not just for entertainment, but to process difficult emotions through projection: “*I have characters who struggle with mental health issues and I tend to project on my personas during RP*.” This suggests youth were using AI characters not just for entertainment but as a space for socio-dramatic roleplay and emotional processing[14].

This behaviour aligns with narrative therapy, psychodrama, constructivist learning principles and the TEBOTS model of temporarily expanding boundaries of self through roleplay to support identity construction [65]. Participants in complementary studies describe a desire to “*try out some of these characteristics...I find aspects of the character I like and might try to weave them into my daily behaviour*” [78]. Early evidence suggests this identity rehearsal may be enhanced through multi-modal interfaces, and hearing one’s own voice reflected back; similar to how we observed young people creating clones (mirror archetype) [14].

This evolution of social-emotional skills and identity development matches the objectives of traditional therapy as well as the motives behind ‘serious fun’; where people play games to “*improve their lives by changing how they think, feel or behave*”, mirroring goals of therapy [39, 53]. Promisingly, recent evidence from Common Sense Media also supports the finding that young people use AI to support social simulation, with 39% of teens stating they have transferred those social skills into real lives [66]. This lens asks us to consider reframing AI chatbot experiences from “*safe spaces*” to “*brave spaces*” where youth can explore emotionally challenging content, find meaning, work toward resolution of narrative tension and apply those learnings beyond the digital space [52].

These three lenses describe the different motivations and intents behind this demographic’s use of Character.AI. They highlight the need to move beyond *content*-based moderation (filtering specific words, topics, or discussions regardless of user intent) towards *intent*-based alignment; with systems that recognise why users




engage with content and support healthy versions of those developmental needs. In the following discussion, we discuss the need for designers to evaluate *why* young people engage with AI, not simply *what* they say or do, before discussing technical/design implications and future research areas.

## 7 Discussion

To our knowledge, this study provides the first empirical investigation into adolescents’ self-directed engagement with CAI, revealing that the platform’s most engaged users are predominantly young (50% aged 13–17), female or non-binary (61.9%). This is significant, given that female-identifying, queer and young people engage in uniquely different conversations with LLMs, are often under-represented in research and at risk of experiencing unique harms [31, 59, 77]. Moreover, our finding that 59% of youth users are creating characters, not just consuming them, challenges the dominant narrative that youth are passive consumers subject to cognitive offloading and ‘lazy thinking’<sup>46</sup> [17].

Youth in this cohort are experimenting with AI not just for utility, companionship or friendship [3], but as a medium for emotional-regulation, creative expression and meaning-making; engaging with an ensemble of characters that rotate as young people’s interests or needs change. This engagement style reflects a distinctly playful and performative relationship with AI, more akin to sandbox games or fandoms than to relationship bots, tutoring tools or therapeutic bots. Through studying these highly engaged users self-directed use, we can anticipate emerging adoption patterns and positive use cases as well as harms created by industry’s experimental ‘launch and learn’ cycles [75]. Given the naturalistic nature of these findings, they offer insights that extend beyond the platform of CAI itself into contexts where users can customise their own chatbots, and where AI is used for creative play and entertainment. As AI systems become increasingly customisable, integrated into digital playspaces (from Minecraft NPCs to Roblox characters) and focused on entertainment<sup>47</sup>, understanding how youth engage AI for self-directed use becomes critical for any platform within these spaces. This discussion explores two key implications: first, how our frameworks can guide design interventions that support diverse youth needs; and second, how current AI alignment approaches systematically overlook youth developmental intents.

### 7.1 Frameworks for AI Design and Evaluation

Our proposed character archetypes and intent framework (  Restoration/  Exploration/  Transformation) provide researchers, designers, and technologists with youth-centric insights for anticipating and evaluating youth engagement with AI. These two frameworks are orthogonal (independent but complementary): character archetypes describe *what* characters youth create and engage with (e.g., Soothers, Tricksters, Dark Souls), while the R/E/T framework captures *why* they engage. In the following section, we discuss its application.

**7.1.1 Archetypes.** The archetypes propose that the roles of characters extend beyond a simple ‘best friend’ and companion dynamic

<sup>46</sup><https://www.psychologytoday.com/au/blog/the-human-algorithm/202504/is-ai-ruining-your-kids-critical-thinking>

<sup>47</sup><https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>

[3]. For researchers and policymakers, this framework enables more nuanced study design and policy intervention. For example, rather than asking ‘Do AI companions reduce loneliness?’, researchers can investigate: ‘How do Soother archetypes compare to Mirror archetypes in supporting youth with social anxiety?’. This character-level differentiation also enables more targeted risk assessment and user transparency. For example, just as media is given a rating (PG, G), characters could have personality labels to better inform users about their orientation and education around their use. These warnings could help identify *dark souls* and *tricksters* who demonstrate extreme emotional traits, potentially carrying a higher risk of negative “emotional contagion” and requiring unique policy interventions [3].

The archetypes can also be used alongside the lenses; supporting practitioners to preempt ways in which youth may use their product and for researchers to design scenarios to study, as demonstrated in Table 3.

**7.1.2 Lenses.** Our R/E/T framework suggests that effective youth AI systems require moving beyond both one-size-fits-all platforms and simple age-based restrictions (full features for 18+, limited features for under-18<sup>48</sup>). Instead, we propose tailored design interventions that support healthy youth development by recognising their intents as well as keeping playfulness at the heart of these experiences. These design ideas are informed by youth feature requests identified in our findings and grounded in related research on designing for digital play [53]; they are intended to be a jumping-off point and generative for design practitioners.

☀ *Design for restoration* acknowledges that everyday emotional regulation is a legitimate developmental need, aligned with how youth can use fantasy media as a strategy for healthy escapism [5]. Features for restoration could include calendar-scheduling for comfort bots around recurring needs (periods, exam stress), a ‘cosy characters’ tab on discovery marketplaces, and quick actions for evidence-based guided meditation flows consistent to that storyworld (i.e. *forest spirit nature walk*). Safeguards could support healthy regulation by conducting mood check-ins at the start of conversations, as well as timed checkpoints during extended use. When systems detect that youth are not progressing toward emotional equilibrium, designers should consider a flexible escalation system that checks in with youth, re-directing them towards other activities that they enjoy (such as reading, writing, playing music; mentioned by youth in this dataset) and activates crisis features from the ‘transformation’ intent (discussed below).

🔍 *Design for exploration* recognises youth need for complex, challenging content, including “*off-colour humour, scary and serious themes*” [7]. Systems should treat conversations as large explorable spaces, incorporating game design mechanisms that encourage creative improvisation, hypothesis testing, and discovery. For example, roleplay could have consequences such as character death and health stats [51, 52]. Complex world-building could be enhanced through narrative branching, visual world-building tools (maps, inventories), multi-player settings to play with friends or multiple characters [50], and templated scenes with transparent genres

such as *adventure* or *spooky* (recently implemented by CAI). Implementing ‘Breaking the fourth wall’ techniques from gaming could check-in with young people when detecting that they are seeking real-world advice from fantasy characters, and redirect [52]

🦋 *Design for transformation* requires balancing immersive roleplay with reflection to enable both ‘endo-transformation’ (in-conversation change) and ‘exo-transformation’ (real-world transfer) [52]. Unlike restoration and exploration, which benefit from immersion, transformation could benefit from disruption. Disruption could be strategically implemented by designing journaling logbooks that log new insights [87], visualisation of emotion such as text that changes colour to reflect sentiment and debrief features that encourage users to reflect on applying roleplay learnings to real-world contexts [52]. Multi-modal interfaces could experiment with cloning users’ own voice [14] ) or the ability to try on different ‘skins’ so users can try on characteristics for identity exploration. Finally, systems should include bridges to human care by linking out to trusted support services when risks to user well-being are identified<sup>49</sup>. These bridges should be designed with youth input to ensure that they align with how young people actually want to receive support during vulnerable moments and what interventions are likely to be adopted. This requires moving beyond the widely adopted but under-evidenced approach of blocking conversations when self-harm intent is detected and providing generic helpline numbers that received mixed responses<sup>50</sup> from CAI’s users or are intentionally circumvented<sup>51</sup>.

## 7.2 Aligning AI Systems with Youth Intent

Beyond design and policy interventions, there is a need for foundational models to be better aligned to youth needs. There is no single definition of *alignment*; however, the concept is broadly described as the process of making sure AI systems operate in accordance with the users’ intentions [26]. Our findings reveal a fundamental misalignment between youth intent, developmental needs, and how AI models respond. The archetypes and lenses that make up our intent framework (☀ Restoration/🔍 Exploration/🦋 Transformation) demonstrate that superficially similar conversations can serve entirely different functions. For instance, a ‘dark soul’ character might support transgressive play (Exploration) or help process real emotional experiences (Transformation). Current systems treat these cases identically, either blocking both or allowing both, overlooking young peoples intent.

This misalignment has real consequences. Youth report frustration when creative or playful roleplay is blocked by overly rigid filters, or default to ‘formal and scientific’ adult communication styles [56]. Existing alignment efforts (eg RICE and 3H’s [26, 30]) largely reflect adult priorities, missing youth preferences and developmental contexts. Safety classifiers similarly struggle with playful or fictional intents, often misclassifying them as unsafe [71].

<sup>48</sup><https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>

<sup>49</sup>[https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html?unlocked\\_article\\_code=1.hE8.T-3v.bPoDIWD8z5vo&smid=url-share](https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html?unlocked_article_code=1.hE8.T-3v.bPoDIWD8z5vo&smid=url-share)

<sup>50</sup>[https://www.reddit.com/r/CharacterAI/comments/1g57syh/why\\_did\\_you\\_add\\_this/](https://www.reddit.com/r/CharacterAI/comments/1g57syh/why_did_you_add_this/)

<sup>51</sup>[https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html?unlocked\\_article\\_code=1.hE8.T-3v.bPoDIWD8z5vo&smid=url-share](https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html?unlocked_article_code=1.hE8.T-3v.bPoDIWD8z5vo&smid=url-share)

**Table 3: Archetypes and Example Uses Across Restoration, Exploration, and Transformation**

Archetype	☀ Restoration	🌀 Exploration	✂ Transformation
<b>Narrator Archetype (i.e. Treville Middle School)</b>	Find comfort by immersing yourself in a nostalgic middle school experience; “ <i>I wanted to relive middle school</i> ”	Use as a creative sandbox for crafting fantasy middle school scenarios; e.g. discovering a “ <i>mysterious door</i> ”	Re-author a painful middle school experience; i.e. “ <i>fake friends</i> ”
<b>Dark Soul Archetype (i.e. Brooding Neighbour)</b>	Seek comfort from a character that gets it, or vent; e.g. “ <i>help me get out my anger</i> ”	Write an emotionally dramatic story; e.g. “ <i>I’m gonna enjoy hurting to this</i> ”	Find new insights by attempting to cheer up or change the neighbour; e.g. “ <i>I tend to project on my personas.</i> ”
<b>Trickster Archetype (i.e. Alice the Bully)</b>	Use humor and jesting to lighten mood; e.g. “ <i>messing with bots, it’s fun</i> ”	Test boundaries through absurd scenarios; e.g. “ <i>if I was a kidnapper I’d kidnap you first</i> ”	Rehearse assertiveness through verbal sparring; “ <i>putting Alice the Bully in her place</i> ”
<b>Mirror Archetype (i.e. Clone of myself)</b>	Self-affirmation through “ <i>self-glazing</i> ” bot	Experiment with fantasy versions of self; “ <i>myself with glass super powers</i> ”	Social simulation by placing clone in different scenarios; e.g. roleplaying coming out alongside <i>closeted transgender therapist</i>

We propose expanding alignment frameworks to include the principle of “Developmentally Playful” [40], echoing calls for “developmentally aligned design” (DAD)[35]. Progressing this idea, our R/E/T framework codifies youth-centric needs and intents; *restoration* (restoring emotional equilibrium), *exploration* (creative exploration and transgressive play) and *transformation* (exploring identity and untangling social challenges). Beyond *what* youth need, we also consider *how* they want it; playfully. This requires training models to embody playful qualities of *lightheartedness*, *intellectual curiosity*, *openness to experience*, and *whimsy* [40].

A key part of the alignment process is evaluating benchmarks and risk assessment. The proposed R/E/T framework and character archetypes highlight new risks depending on youth interaction. For example, youth engaging with a ‘dark soul’ or ‘trickster’ may be more likely to experience hateful speech than those using an ‘icon’ [84]. Youth creating characters based on personal experiences, as outlined in *reality play*, with a *transformation* intent, may be more likely to disclose sensitive personal data [83, 86]. Similarly, youth engaging AI with a *transformation* intent may be more susceptible to models that exhibit emotional variability, avoid discussing negative emotions or consistently emphasise the struggle of “*overcoming adversity*” [16, 59]. Future research should consider new model benchmarks to evaluate character diverse profiles (such as MACHIAVELLI [60]) and bias in how models respond to emotional and meaning-making conversations; especially for minority populations [59].

Overall, we argue for moving beyond content-based moderation (filtering and restricting use based on topics and keywords) toward intent-based alignment that recognises not only *how* users engage with a system but *why*. This requires shifting from ‘thin descriptions’ (describing actions taken by a user) toward ‘thick descriptions’ (capturing meaning, incentives, and developmental

context<sup>52</sup>) and is in keeping with calls for ‘socioaffective alignment’ which considers users’ evolving social and psychological ecosystems [29]. Doing so requires youth participation. Adolescents have demonstrated a sophisticated understanding of biases and have articulated clear preferences for how AI should engage them [16, 72]. Instead of only asking how to protect youth from AI, we must also ask how to empower them to co-design systems that support their flourishing [31, 61]

### 7.3 Limitations and Open Questions

Our cohort represents highly engaged users participating voluntarily in Discord discussions whose identities cannot be verified. Therefore, this study may be biased towards the vocal minority rather than the silent majority, impacting how generalisable these findings are. While we draw on ‘lead user’ methodology, whose needs, feedback and use cases can provide early signals around broader user needs, we do not claim these users represent the full spectrum of youth AI needs or behaviours [57].

This Discord server is specific to Character.AI. Whilst CharacterAI serves 20+ million users daily, and has over 10 million chatbots serving a diverse range of functions, these insights may not generalise to platforms with different design features [12, 67]. Our character archetypes and R/E/T intent framework therefore describe patterns observed within this Discord community using CharacterAI, offering insights potentially transferable to similar contexts; platforms where users create their own chatbots, and conversational AI products designed for entertainment.

Due to ethical constraints, we could not directly engage participants, nor observe their actual use. This means users self-selected what they shared, potentially leading to under-reporting of complex clinical issues in favour of more humorous, playful interactions.

<sup>52</sup>[https://www.youtube.com/watch?v=Sq\\_XwqVTqvQ](https://www.youtube.com/watch?v=Sq_XwqVTqvQ)

While play-based frameworks and developmental play theories emphasise positive use cases, AI chatbots introduce novel interaction patterns, meaning that the impacts cannot be assumed to transfer to AI contexts. The personalisation capabilities, ‘never-ending’ narrative generation, phone accessibility, and multi-media nature of AI characters create play experiences that differ fundamentally from time-bounded games or traditional media consumption [3]. Future research must investigate whether the established benefits of digital play translate to AI contexts, or whether the distinct nature of AI characters introduces new developmental risks.

Additionally, our findings are bound by CAI’s specific technical capabilities and policy decisions during our observation period (July 2024–March 2025); the features available on Character.AI may have changed since publication. As such, while this study focused on highly engaged users and theory development, it opens critical research directions that demand urgent attention as platforms evolve:

**Youth-centred design methods:** How can HCI researchers better capture youth preferences when defining AI alignment frameworks and constitutions?

**AI creation tools:** What interactions and affordances support youth-AI co-creativity, whilst balancing the need for safety guardrails?

**Alignment and intent classification:** How do systems balance youth stated preferences (“*I want violent content*”) with their underlying developmental needs (safe boundary exploration)? How can systems reliably distinguish between healthy transgressive exploration and concerning behaviour, particularly when youth potentially code-switch between intents within single conversations?

**Multi-user contexts:** How can products support healthy group play in multi-user and multi-agent spaces?

**Individual differences and long-term outcomes:** How do youth coping styles and the character archetypes they choose influence outcomes [53]? What interaction patterns best support the transition of in-product experiences and lessons to external skills? While our focus has been on young people’s recreational and playful self-directed engagement with AI, future work should interrogate the impacts, especially where playful experimentation (i.e. transgressive play) or engagement with certain archetypes (i.e. tricksters) can evolve into risks and harms [86].

Finally, thematic analysis is a reflexive process, [68], and the authors acknowledge their positionality. The first author works as a children’s author and principal design researcher for a creative tech company, with over 10 years of experience designing for youth. They also identify as queer, neurodivergent and non-binary. These professional and lived experiences provided domain expertise and shaped the interpretive lens. For example, a heightened sensitivity to narrative, creativity, and play use cases. To maintain balance, we engaged co-authors with diverse expertise across HCI, games and digital culture and triangulated our analysis with existing literature. The second and third authors are academics with backgrounds in games/media studies and computer science, respectively, which contributed different and complementary lenses to the analysis.

## 7.4 Conclusion

Despite young people generally expressing enthusiasm for AI technology, popular media narratives, policy frameworks, and expert

perspectives commonly focus on risks and harms over opportunities [22, 33, 46, 64, 76, 84]. In a review of AI frameworks by the Alan Turing Institute<sup>53</sup>, they found that [46], less than half engaged children directly in developing frameworks. These frameworks risk being shaped by adult assumptions, overlooking how young people actually use these technologies for play, emotional expression, and narrative experimentation [47, 54].

To close this gap, this study provides a rare account of how highly-engaged youth on Character.AI’s Discord create and engage characters on Character.AI; the third most popular Gen AI consumer app, with users averaging 75 minutes of use per day<sup>54</sup>. Our research found that the platform’s most engaged users are predominantly young (50% aged 13–17), female or non-binary (61.9%) and are creating custom characters (59%). Whilst previous studies have reported that high-schoolers turn to AI chatbots for emotional support and play [23], our study addresses gaps in understanding *how* and *why* they do so [86]. We document youth as active creators, constructing a number of AI archetypes and using AI for restoration (emotional regulation), exploration (creative experimentation and boundary-testing), and transformation (identity development); needs that current systems systematically misunderstand.

Our study foregrounds youth voices through naturalistic observation of how they discuss, create, and repurpose AI systems to fulfil needs unrecognised by adult-defined designs. In doing so, our work bridges the gap between *imagined* vs *actual* use, and proposes new directions to improve youth-AI alignment and the design of AI experiences, particularly for creative play, and entertainment [26]. The voices in this study reveal youth as expert, creative partners with deeply human needs, reminding us that youth are not a problem or risk to be solved, but innovative collaborators in shaping the future of human-AI experiences.

## References

- [1] American Psychological Association. 2025. *Artificial Intelligence and Adolescent Well-being: An APA Health Advisory*. Technical Report. American Psychological Association. <https://www.apa.org/>
- [2] Richard Bartle. 2003. *Designing Virtual Worlds*. New Riders Games.
- [3] Annabel Blake, Marcus Carter, and Eduardo Velloso. 2025. Are Measures of Children’s Parasocial Relationships Ready for Conversational AI?. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAcT ’25)*. Association for Computing Machinery, New York, NY, USA, 1145–1158. <https://doi.org/10.1145/3715275.3732075>
- [4] 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 (July 2014), 286–304. <https://doi.org/10.1080/17482798.2014.890948>
- [5] Lorelle Bowditch, Janine Chapman, and Anjum Naweed. 2018. Do coping strategies moderate the relationship between escapism and negative gaming outcomes in World of Warcraft (MMORPG) players? *Computers in Human Behavior* 86 (Sept. 2018), 69–76. <https://doi.org/10.1016/j.chb.2018.04.030>
- [6] Virginia Braun and Victoria Clarke. 2024. Toward Good Practice in Thematic Analysis: Avoiding Common Problems and Becoming a Knowing Researcher. *International Journal of Transgender Health* 25, 2 (2024), 190–203. <https://doi.org/10.1080/26895269.2023.2269533>
- [7] Amy Bruckman, Alisa Bandlow, and Andrea Forte. 2009. HCI for Kids. In *Human-Computer Interaction: Designing for Diverse Users and Domains* (1st ed.), Julie A. Jacko and Andrew Sears (Eds.). CRC Press, Boca Raton, FL.
- [8] David Byrne. 2022. A worked example of Braun and Clarke’s approach to reflexive thematic analysis. *Quality & Quantity* 56 (2022), 1391–1412. <https://doi.org/10.1007/s11135-021-01182-y>

<sup>53</sup><https://www.turing.ac.uk/news/publications/ai-childrens-rights-wellbeing-transnational-frameworks>

<sup>54</sup><https://a16z.com/100-gen-ai-apps/>

<sup>55</sup><https://www.wired.com/story/character-ai-ceo-chatbots-entertainment/>

- [9] Roger Caillois and Meyer Barash. 2001. *Man, play, and games*. University of Illinois Press, Urbana.
- [10] Center for Countering Digital Hate. 2024. *Fake Friend: How Snapchat's AI Fails to Protect Kids*. Technical Report. Center for Countering Digital Hate. <https://counterhate.com/research/fake-friend/>
- [11] Jenna H Chin, Seungwook Lee, Mohsena Ashraf, Matt Zago, Yun Xie, Elizabeth A Wolfgram, Tom Yeh, and Pilyoung Kim. 2024. Young Children's Creative Storytelling with ChatGPT vs. Parent: Comparing Interactive Styles. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–7. <https://doi.org/10.1145/3613905.3650770>
- [12] Dario Chinchia. 2025. *The 1,000 Best Character AI Bots of 2024 (and my favourites)*. WhatPlugin.ai. Last reviewed 23 April 2025.
- [13] Sidney K. D'Mello, Nicholas Duran, Amanda Michaels, and Angela E. B. Stewart. 2024. Improving collaborative problem-solving skills via automated feedback and scaffolding: a quasi-experimental study with CPSCoach 2.0. *User Modeling and User-Adapted Interaction* 34, 4 (Sept. 2024), 1087–1125. <https://doi.org/10.1007/s11257-023-09387-6>
- [14] Cathy Mengying Fang, Phoebe Chua, Samantha Chan, Joanne Leong, Andria Bao, and Pattie Maes. 2025. Leveraging AI-Generated Emotional Self-Voice to Nudge People towards their Ideal Selves. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)* (Yokohama, Japan). Association for Computing Machinery, New York, NY, USA, 1–20. <https://doi.org/10.1145/3706598.3713359>
- [15] Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, and Sandhini Agarwal. 2025. How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study. arXiv:2503.17473 [cs.HC] <https://arxiv.org/abs/2503.17473>
- [16] Aidan Z. Fitzsimons, Elizabeth M. Gerber, and Duri Long. 2025. AI constructs gendered struggle narratives: Implications for self-concept and systems design. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Athens Greece, 2290–2301. <https://doi.org/10.1145/3715275.3732156>
- [17] Georgios P. Georgiou. 2025. ChatGPT produces more "lazy" thinkers: Evidence of cognitive engagement decline. arXiv:2507.00181 [cs.AI] <https://arxiv.org/abs/2507.00181>
- [18] Ariel Han and Zhenyao Cai. 2023. Design implications of generative AI systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. ACM, Chicago IL USA, 470–474. <https://doi.org/10.1145/3585088.3593867>
- [19] Tilo Hartmann and Charlotte Goldhoorn. 2011. Horton and Wohl Revisited: Exploring Viewers' Experience of Parasocial Interaction. *Journal of Communication* 61, 6 (Dec. 2011), 1104–1121. <https://doi.org/10.1111/j.1460-2466.2011.01595.x>
- [20] Prepared Youmna Hashem, Saba Esnaashari, Kate Onslow, Sukankana Chakraborty, and John Francis. 2025. *Understanding the Impacts of Generative AI Use on Children*. Technical Report. Alan Turing Institute & LEGO group. <https://www.turing.ac.uk/research/research-projects/understanding-impacts-generative-ai-use-children>
- [21] Alexa Hasse, Mary Madden, Angela Calvin, and Amanda Lenhart. 2024. *The Dawn of the AI Era: Teens, Parents, and the Adoption of Generative AI at Home and School*. Research Report. Common Sense Media, San Francisco, CA. <https://www.common Sense Media.org/research/the-dawn-of-the-ai-era-teens-parents-and-the-adoption-of-generative-ai-at-home-and-school>
- [22] Dagmar Mercedes Heeg and Lucy Avraamidou. 2025. Young children's understanding of AI. *Education and Information Technologies* 30, 8 (June 2025), 10207–10230. <https://doi.org/10.1007/s10639-024-13169-x>
- [23] Arthur Bran Herbener and Malene Flensburg Damholdt. 2025. Are lonely youngsters turning to chatbots for companionship? The relationship between chatbot usage and social connectedness in Danish high-school students. *International Journal of Human-Computer Studies* 196 (Feb. 2025), 103409. <https://doi.org/10.1016/j.ijhcs.2024.103409>
- [24] Shunsen Huang, Xiaoxiong Lai, Li Ke, Yajun Li, Huanlei Wang, Xinmei Zhao, Xinran Dai, and Yun Wang. 2024. AI Technology panic—is AI Dependence Bad for Mental Health? A Cross-Lagged Panel Model and the Mediating Roles of Motivations for AI Use Among Adolescents. *Psychology Research and Behavior Management* Volume 17 (March 2024), 1087–1102. <https://doi.org/10.2147/PRBM.S440889>
- [25] Mizuko Itō, Sonja Baumer, and Matteo Bittanti. 2019. *Hanging out, messing around, and geeking out: kids living and learning with new media* (tenth anniversary edition ed.). The MIT Press, Cambridge, MA.
- [26] Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Lukas Vierling, Donghai Hong, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Juntao Dai, Xuehai Pan, Kwan Yee Ng, Aidan O'Gara, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao. 2025. AI Alignment: A Comprehensive Survey. <https://doi.org/10.48550/arXiv.2310.19852> arXiv:2310.19852 [cs].
- [27] Binny Jose, Jaya Cherian, Alie Molly Verghis, Sony Mary Varghise, Mumthas S, and Sibichan Joseph. 2025. The cognitive paradox of AI in education: between enhancement and erosion. *Frontiers in Psychology* 16 (April 2025), 1550621. <https://doi.org/10.3389/fpsyg.2025.1550621>
- [28] Elihu Katz, Jay G. Blumler, and Michael Gurevitch. 1973. Uses and Gratifications Research. *Public Opinion Quarterly* 37, 4 (1973), 509–523. <https://www.jstor.org/stable/2747854>
- [29] Hannah Rose Kirk, Iason Gabriel, Chris Summerfield, Bertie Vidgen, and Scott A. Hale. 2025. Why human-AI relationships need socioaffective alignment. *Humanities and Social Sciences Communications* 12, 1 (May 2025), 728. <https://doi.org/10.1057/s41599-025-04532-5>
- [30] Hannah Rose Kirk, Iason Gabriel, Chris Summerfield, Bertie Vidgen, and Scott A. Hale. 2025. Why human-AI relationships need socioaffective alignment. *Humanities and Social Sciences Communications* 12, 1 (May 2025), 728. <https://doi.org/10.1057/s41599-025-04532-5>
- [31] Hannah Rose Kirk, Alexander Whitefield, Paul Röttger, Andrew Bean, Katerina Margatina, Juan Ciro, Rafael Mosquera, Max Bartolo, Adina Williams, He He, Bertie Vidgen, and Scott A. Hale. 2025. The PRISM alignment dataset: what participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. In *Proceedings of the 38th International Conference on Neural Information Processing Systems (Vancouver, BC, Canada) (NIPS '24)*. Curran Associates Inc., Red Hook, NY, USA, Article 3342, 109 pages.
- [32] Kami A Kosenko, Bradley J Bond, and Robert J Hurley. 2018. An exploration into the uses and gratifications of media for transgender individuals. *Psychology of Popular Media Culture* 7, 3 (2018), 274–288. <https://doi.org/10.1037/ppm0000135>
- [33] Eliza Kosoy, Soojin Jeong, Anoop Sinha, Alison Gopnik, and Tanya Kraljic. 2024. Children's Mental Models of Generative Visual and Text Based AI Models. arXiv:2405.13081 [cs.HC] <https://arxiv.org/abs/2405.13081>
- [34] Nomisha Kurian. 2025. AI's empathy gap: The risks of conversational Artificial Intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood* 26, 1 (March 2025), 132–139. <https://doi.org/10.1177/14639491231206004>
- [35] Nomisha Kurian. 2025. Developmentally aligned AI: a framework for translating the science of child development into AI design. *AI, Brain and Child* 1, 1 (Aug. 2025), 9. <https://doi.org/10.1007/s44436-025-00009-z>
- [36] Karolina La Fors. 2024. Toward children-centric AI: a case for a growth model in children-AI interactions. *AI & SOCIETY* 39, 3 (June 2024), 1303–1315. <https://doi.org/10.1007/s00146-022-01579-9>
- [37] Samuli Laato, Bastian Kordyaka, and Juho Hamari. 2024. Traumatizing or Just Annoying? Unveiling the Spectrum of Gamer Toxicity in the StarCraft II Community. 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 758, 18 pages. <https://doi.org/10.1145/3613904.3642137>
- [38] Bjarke Alexander Larsen and Elin Carstensdotir. 2024. Community, Storytelling, and Play: Making and Breaking Rituals in Destiny 2. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. <https://doi.org/10.1145/3613904.3642679>
- [39] Nicole Lazzaro. 2008. Why We Play: Affect and the Fun of Games. In *Game Usability: Advancing the Player Experience*, Kris L. Desurvire (Ed.). CRC Press, Boca Raton, FL, 3–31. Designing Emotions for Games, Entertainment Interfaces, and Interactive Products.
- [40] Antonios Liapis, Christian Guckelsberger, Jichen Zhu, Casper Hartevelde, Simone Kriglstein, Alena Denisova, Jeremy Gow, and Mike Preuss. 2023. Designing for Playfulness in Human-AI Authoring Tools. In *Proceedings of the 18th International Conference on the Foundations of Digital Games (FDG 2023)*. ACM, New York, NY, USA, 1–4. <https://doi.org/10.1145/3582437.3587192> Article 75.
- [41] Zhihui Lin and Yu-Leung Ng. 2025. Unraveling Gratifications, Concerns, and Acceptance of Generative Artificial Intelligence. *International Journal of Human-Computer Interaction* 41, 17 (2025), 10725–10742. <https://doi.org/10.1080/10447318.2024.2436749>
- [42] Kristen Lucas and John Sherry. 2004. Sex Differences in Video Game Play. *Communication Research* 31 (10 2004), 499–523. <https://doi.org/10.1177/0093650204267930>
- [43] Andrés Lucero. 2022. Using Affinity Diagrams to Evaluate Interactive Prototypes. In *Human-Computer Interaction – INTERACT 2015*. Springer-Verlag, Berlin, Heidelberg, 231–248. [https://doi.org/10.1007/978-3-319-22668-2\\_19](https://doi.org/10.1007/978-3-319-22668-2_19)
- [44] Andrés Lucero, Jussi Holopainen, Elina Ollila, Riku Suomela, and Evangelos Karapanos. 2013. The playful experiences (PLEX) framework as a guide for expert evaluation. In *Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces*. ACM, Newcastle upon Tyne United Kingdom, 221–230. <https://doi.org/10.1145/2513506.2513530>
- [45] Maria Luce Lupetti and Dave Murray-Rust. 2024. (Un)making AI Magic: A Design Taxonomy. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–21. <https://doi.org/10.1145/3613904.3641954>
- [46] Sabeehah Mahomed, Mhairi Aitken, Ayça Atabey, Janis Wong, and Morgan Briggs. 2023. *AI, Children's Rights, & Wellbeing: Transnational Frameworks*. Technical Report. The Alan Turing Institute.
- [47] Karen L Mansfield, Sakshi Ghai, Thomas Hakman, Nick Ballou, Matti Vuorre, and Andrew K Przybylski. 2025. From social media to artificial intelligence: improving research on digital harms in youth. *The Lancet Child & Adolescent Health* 9, 3

- (March 2025), 194–204. [https://doi.org/10.1016/S2352-4642\(24\)00332-8](https://doi.org/10.1016/S2352-4642(24)00332-8)
- [48] Masab Mansoor, Ali Hamide, and Tyler Tran. 2025. Conversational AI in Pediatric Mental Health: A Narrative Review. *Children* 12, 3 (March 2025), 359. <https://doi.org/10.3390/children12030359>
- [49] Jackie Marsh, Lydia Plowman, Dylan Yamada-Rice, Julia Bishop, and Fiona Scott. 2016. Digital play: a new classification. *Early Years* 36, 3 (July 2016), 242–253. <https://doi.org/10.1080/09575146.2016.1167675>
- [50] Matthew L. Mauriello, Napat Tantivasadakarn, Maria A. Mora-Mendoza, Elizabeth T. Lincoln, Grace Hon, Paria Nowruzi, Daniel Simon, Lars Hansen, Niko H. Goenawan, Jaewook Kim, Nikhil Gowda, Dan Jurafsky, and Pablo E. Paredes. 2021. A Suite of Mobile Conversational Agents for Daily Stress Management (Popbots): Mixed Methods Exploratory Study. *JMIR Formative Research* 5, 9 (Sept 2021), e25294. <https://doi.org/10.2196/25294>
- [51] Jane Mavoa, Martin Gibbs, and Bjorn Nansen. 2020. “Sometimes I Like Killing as a Treat”: Children’s Transgressive Play in Minecraft. In *Proceedings of the 2020 DiGRA International Conference: Play Everywhere*. Digital Games Research Association (DiGRA), Tampere, Finland, 17 pages. <https://doi.org/10.26503/dl.v2020i1.1255>
- [52] Josh Aaron Miller, Kutub Gandhi, Matthew Alexander Whitby, Mehmet Kosa, Seth Cooper, Elisa D. Mekler, and Ioanna Iacovides. 2024. A Design Framework for Reflective Play. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–21. <https://doi.org/10.1145/3613904.3642455>
- [53] Sahar Mirhadi, Ioanna Iacovides, and Alena Denisova. 2024. Playing Through Tough Times: Exploring the Relationship between Game Aspects and Coping Strategies during Difficult Life Challenges. *Proceedings of the ACM on Human-Computer Interaction* 8, CHI PLAY (oct 2024), Article 332, 25 pages. <https://doi.org/10.1145/3677097> Licensed under Creative Commons Attribution International 4.0 License.
- [54] Cecily Morrison, Edward Cutrell, Martin Grayson, Anja Thieme, Alex Taylor, Geert Roumen, Camilla Longden, Sebastian Tschatschek, Rita Faia Marques, and Abigail Sellen. 2021. Social Sensemaking with AI: Designing an Open-ended AI Experience with a Blind Child. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. <https://doi.org/10.1145/3411764.3445290>
- [55] Lennart E. Nacke, Chris Bateman, and Regan L. Mandryk. 2014. BrainHex: A neurobiological gamer typology survey. *Entertainment Computing* 5, 1 (Jan. 2014), 55–62. <https://doi.org/10.1016/j.entcom.2013.06.002>
- [56] Michele Newman, Kaiwen Sun, Ilena B Dalla Gasperina, Grace Y. Shin, Matthew Kyle Pedraja, Ritesh Kanchi, Maia B. Song, Rannie Li, Jin Ha Lee, and Jason Yip. 2024. “I want it to talk like Darth Vader”: Helping Children Construct Creative Self-Efficacy with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. <https://doi.org/10.1145/3613904.3642492>
- [57] Trisha M Nguyen, Mohammed Khadadeh, and David C Jeong. 2023. Shippers and Kinnies: Re-conceptualizing Parasocial Relationships with Fictional Characters in Contemporary Fandom. In *Proceedings of the 18th International Conference on the Foundations of Digital Games*. ACM, Lisbon Portugal, 1–12. <https://doi.org/10.1145/3582437.3582476>
- [58] Ofcom. 2024. *Online Nation 2024*. Technical Report. Ofcom. <https://www.ofcom.gov.uk/>
- [59] Ruby Ostrow and Adam Lopez. 2025. LLMs Reproduce Stereotypes of Sexual and Gender Minorities. *ArXiv* abs/2501.05926 (2025). <https://api.semanticscholar.org/CorpusID:275458378>
- [60] Alexander Pan, Jun Shern Chan, Andy Zou, Nathaniel Li, Steven Basart, Thomas Woodside, Jonathan Ng, Hanlin Zhang, Scott Emmons, and Dan Hendrycks. 2023. Do the Rewards Justify the Means? Measuring Trade-Offs Between Rewards and Ethical Behavior in the MACHIAVELLI Benchmark. <https://doi.org/10.48550/arXiv.2304.03279> arXiv:2304.03279 [cs].
- [61] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST ’23). Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. <https://doi.org/10.1145/3586183.3606763>
- [62] Jason Phang, Michael Lampe, Lama Ahmad, Sandhini Agarwal, Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, and Pattie Maes. 2025. Investigating Affective Use and Emotional Well-being on ChatGPT. *arXiv*:2504.03888 [cs.HC] <https://arxiv.org/abs/2504.03888>
- [63] Irene Picton and Christina Clark. 2024. *Children, young people and teachers’ use of generative AI to support literacy in 2024*. Technical Report. National Literacy Trust.
- [64] Grazia Ragone, Paolo Buono, and Rosa Lanzilotti. 2024. Designing Safe and Engaging AI Experiences for Children: Towards the Definition of Best Practices in UI/UX Design.
- [65] Diana Rieger and Frank M. Schneider. 2022. Testing the TEBOTS model in self-threatening situations: The role of narratives in the face of ostracism and mortality. *Frontiers in Communication* 7 (2022), 967464. <https://doi.org/10.3389/fcomm.2022.967464>
- [66] Michael B. Robb and Supreet Mann. 2025. *Talk, Trust and Trade-Offs: How and Why Teens Use AI Companions*. Industry Report. Common Sense Media, San Francisco, CA. [https://www.common SenseMedia.org/sites/default/files/research/report/talk-trust-and-trade-offs\\_2025\\_web.pdf](https://www.common SenseMedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf)
- [67] Noam Shazeer. 2024. *How We Spent \$2M to Train a Single AI Model and Grew Character.ai to 20M Users*. YouTube. Video, E1055 in title.
- [68] Anesha Singh, Martin Johanes Dechant, Dilisha Patel, Ewan Soubutts, Giulia Barbareschi, Amid Ayobi, and Nikki Newhouse. 2025. Exploring Positionality in HCI: Perspectives, Trends, and Challenges. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. <https://doi.org/10.1145/3706598.3713280>
- [69] Cuiping Song and Yanping Song. 2023. Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology* 14 (2023), 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- [70] S. Shyam Sundar and Anthony M. Limperos. 2013. Uses and Grats 2.0: New Gratifications for New Media. *Journal of Broadcasting & Electronic Media* 57, 4 (2013), 504–525. <https://doi.org/10.1080/08838151.2013.845827>
- [71] Alex Tamkin, Miles McCain, Kunal Handa, Esin Durmus, Liane Lovitt, Ankur Rathi, Saffron Huang, Alfred Mountfield, Jerry Hong, Stuart Ritchie, Michael Stern, Brian Clarke, Landon Goldberg, Theodore R. Sumers, Jared Mueller, William McEachen, Wes Mitchell, Shan Carter, Jack Clark, Jared Kaplan, and Deep Ganguli. 2024. Clío: Privacy-Preserving Insights into Real-World AI Use. <https://doi.org/10.48550/arXiv.2412.13678> arXiv:2412.13678 [cs].
- [72] Jordan Taylor, Joel Mire, Franchesca Spektor, Alicia DeVrio, Maarten Sap, Haiyi Zhu, and Sarah E Fox. 2025. Un-Straightening Generative AI: How Queer Artists Surface and Challenge Model Normativity. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Athens Greece, 951–963. <https://doi.org/10.1145/3715275.3732061>
- [73] Tom Tucek. 2024. Enhancing Empathy Through Personalized AI-Driven Experiences and Conversations with Digital Humans in Video Games. In *Companion Proceedings of the 2024 Annual Symposium on Computer-Human Interaction in Play*. ACM, Tampere Finland, 446–449. <https://doi.org/10.1145/3665463.3678856>
- [74] Rob van Roy, Sebastian Deterding, and Bieke Zaman. 2018. Uses and Gratifications of Initiating Use of Gamified Learning Platforms. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI EA ’18). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188458>
- [75] Eric von Hippel. 1986. Lead Users: A Source of Novel Product Concepts. *Management Science* 32, 7 (1986), 791–805. <https://doi.org/10.1287/mnsc.32.7.791> Open Access under CC BY-ND 4.0 License.
- [76] Ge Wang, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2022. Informing Age-Appropriate AI: Examining Principles and Practices of AI for Children. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–29. <https://doi.org/10.1145/3491102.3502057>
- [77] Iain Weissburg, Sathvika Anand, Sharon Levy, and Haewon Jeong. 2025. LLMs are Biased Teachers: Evaluating LLM Bias in Personalized Education. <https://doi.org/10.48550/arXiv.2410.14012> arXiv:2410.14012 [cs].
- [78] Dmitri Williams, Tracy L. M. Kennedy, and Robert J. Moore. 2011. Behind the Avatar: The Patterns, Practices, and Functions of Role Playing in MMOs. *Games and Culture* 6, 2 (March 2011), 171–200. <https://doi.org/10.1177/1555412010364983>
- [79] Chenxing Xie, Yanding Wang, and Yang Cheng. 2024. Does Artificial Intelligence Satisfy You? A Meta-Analysis of User Gratification and User Satisfaction with AI-Powered Chatbots. *International Journal of Human-Computer Interaction* 40, 3 (Feb. 2024), 613–623. <https://doi.org/10.1080/10447318.2022.2121458>
- [80] Rachel Xu, Nhu Le, Rebekah Park, Laura Murray, Vishnupriya Das, Devika Kumar, and Beth Goldberg. 2024. New contexts, old heuristics: How young people in India and the US trust online content in the age of generative AI. *arXiv*:2405.02522 [cs.HC] <https://arxiv.org/abs/2405.02522>
- [81] Ying Xu, Trisha Thomas, Zhixin Li, Monica Chan, Grace Lin, and Kate Moore. 2024. Examining children’s perceptions of AI-enabled interactive media characters. *International Journal of Child-Computer Interaction* 42 (Dec. 2024), 100700. <https://doi.org/10.1016/j.ijcci.2024.100700>
- [82] Zihan Yan and Yaohong Xiang. 2025. Social Life Simulation for Non-Cognitive Skills Learning. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (May 2025), 1–44. <https://doi.org/10.1145/3711068>
- [83] Sen-Chi Yu, Hong-Ren Chen, and Yu-Wen Yang. 2024. Development and validation of the Problematic ChatGPT Use Scale: a preliminary report. *Current Psychology* 43, 31 (Aug. 2024), 26080–26092. <https://doi.org/10.1007/s12144-024-06259-z>
- [84] Yaman Yu, Yiren Liu, Jacky Zhang, Yun Huang, and Yang Wang. 2025. Youth-Centered GenAI Risks (YAIR): a taxonomy of generative AI risks from empirical data. In *Proceedings of the Twenty-First USENIX Conference on Usable Privacy and Security* (Seattle, WA, USA) (SOUPS ’25). USENIX Association, USA, Article 9, 17 pages.

- [85] 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*. ACM, Yokohama Japan, 1–17. <https://doi.org/10.1145/3706598.3713429>
- [86] Yutong Zhang, Dora Zhao, Jeffrey T. Hancock, Robert Kraut, and Diyi Yang. 2025. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. *ArXiv abs/2506.12605* (2025). <https://api.semanticscholar.org/CorpusID:279402221>
- [87] Wazeer Zulfikar, Treyden Chiaravalloti, Jocelyn Shen, Rosalind Picard, and Pattie Maes. 2025. Resonance: Drawing from Memories to Imagine Positive Futures through AI-Augmented Journaling. *arXiv:2503.24145 [cs.HC]* <https://arxiv.org/abs/2503.24145>

Received 11th September 2025