

CHARLES UNIVERSITY

Univerzita Karlova

FACULTY OF ARTS

Filozofická fakulta

Institute of Information Studies and Librarianship | New Media Studies

Ústav informačních studií a knihovnictví | Studia nových médií

Master's Thesis

Diplomová práce

Bc. et Bc. David Laufer

AI love you. Gender and intimacy in user content regarding AI chatbot characters from Character.ai

Miluješ mě, AI tebe. Gender a intimita v uživatelském obsahu týkajícím se AI
chatbotových postav z Character.ai

Supervisor | Vedoucí práce: Mgr. Josef Šlerka, Ph.D.

Consultant | Konzultant: Mgr. Ondřej Trhoň

Prague, 2024

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

1. Při psaní této práce jsem použil automatické funkce kontroly pravopisu v Microsoft Word za účelem kontroly pravopisu. Text jsem posléze zkontroloval a dle potřeby upravil a přebírám plnou zodpovědnost za obsah této publikace.

V Praze 15.12.2024

David Laufer

Declaration

I hereby declare that I have compiled this thesis independently, properly listed all used literature and resources, and that my thesis has not been used within another degree program or to gain any other academic title. I fully agree to my work being used for study and scientific purposes.

1. During the preparation of this thesis, the author used Microsoft Word's default automatic spell check functions in order to check spelling. After using this tool, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication.

In Prague on 15.12.2024

David Laufer

Poděkování

Všem, kteří nějakým dílem přispěli k byť jen seabeméně klidnějšímu psaní této práce. Konzultantovi Mgr. Ondřeji Trhoňovi, že se jí ujal a kladně ji směroval. A vedoucímu Mgr. Josefu Šlerkovi, Ph.D., za formální záštitu.

Acknowledgement

Of all those who contributed, however much or little, to a more peaceful writing of this thesis. Of my consultant Mgr. Ondřej Trhoň, who took it up and offered positive directions. And of Mgr. Josef Šlerka, Ph.D., for formal coverage.

Abstract

My thesis contributes to research on Intimacy and Gender within human-chatbot relationships (HCR) and role-play with a focus on the under-researched, but popular, Character.ai platform. I analyze the discourse in top user threads on the r/CharacterAI subreddit and the content of top user-generated intimate partner chatbots on Character.ai, expanding the existing research focused mostly on stereotypically feminized AI into other gender expressions. I uncover strong notions of control – from the platform, users, and chatbots. The top analyzed characters were explicitly dominant, toxic, or overprotective – traits interestingly prominent across genders and sexualities. Cold, possessive, hegemonic masculine chatbots were most popular, followed by rude or contemptuous feminine chatbots among my analyzed characters. On Reddit, the use of such masculine chatbots was, however, discussed often in relation to user shame or critique and feminine chatbots were almost exclusively discussed as caring, or instead cared for – contrasting my analyses. Similarly, queering approaches, critique or subversions of the gender hegemony in both chatbot and user masculinities, femininities, and non-binary identities constituted the top category of the subreddit discourse on Gender but were mostly reflected in less popular chatbots on Character.ai. As such, my work found discursive intimacy and gender construction also between users, together with their stances on loneliness, addiction, AI, data, or disclosure towards chatbots vs. humans. My analyses are situated within a theoretical framework of algorithmic intimacies, parasocial relationships, gender studies and queer theory, enriching the fields of human-chatbot relationships and critical studies of artificial intelligence.

Abstrakt

Má práce přispívá k výzkumu intimacy a genderu ve vztazích mezi lidmi a chatboty se zaměřením na málo prozkoumanou, ale populární platformu Character.ai. Analyzuji diskurz v nejpopulárnějších uživatelských vláknech na subredditu r/CharacterAI a obsah nejpopulárnějších intimních partnerských chatbotů generovaných uživatelstvem na Character.ai. Rozšiřuji tak existující výzkum zaměřený převážně na stereotypizovaně feminizovanou AI i do oblasti ostatních genderových vyjádření. Odhaluji výrazný prvek kontroly – od platformy, uživatelstva i chatbotů. Nejpopulárnější analyzovaní chatboti byli dominantní, toxici nebo přehnaně ochranní – rysy pozoruhodně často prominentní napříč genderem i sexualitou. Chladní, majetníci, hegemonicky maskulinní chatboti byli

nejpopulárnější, následování drzími či pohrdavými feminními chatboty. Na Redditu byla preference této maskulinity ale diskutována převážně kriticky či v souvislosti se studem, přičemž chatbotová feminita byla téměř výhradně diskutována jako pečující, či naopak opečovávaná – což odlišuje mé dvě analýzy. Přístupy kvíření a subverze genderové hegemonie, odhalující škálu maskulinit, feminit a nebinárních identit u uživatelstva i AI postav, byly pak nejvýraznějším tématem v subredditovém diskurzu o genderu, reflektovaly je nicméně spíše méně populární chatboti na Character.ai. Nacházím tak diskurzivní povahu intimity a genderové konstrukce i mezi uživatelstvem spolu s jeho postoji ohledně AI, samoty, závislosti, dat či svěřování se chatbotům oproti lidem. Mé analýzy jsou zasazeny do teoretického rámce algoritmických intimit, parasociálních vztahů, genderových studií i queer teorie a obohacují pole vztahů člověk-chatbot a kritických studií umělé inteligence.

Keywords

chatbot, Character.ai, gender, intimacy, human-chatbot relationship, AI girlfriend, AI boyfriend, masculinity, femininity, social media, Reddit, artificial intelligence

Klíčová slova

chatbot, Character.ai, gender, intimita, vztah člověk-chatbot, AI přítelkyně, AI přítel, maskulinita, feminita, sociální sítě, Reddit, umělá inteligence

Reference

Laufer, D. (2024). *AI love you. Gender and intimacy in user content regarding AI chatbot characters from Character.ai*. Prague. 147 p. Master's thesis (Mgr.). Charles University, Faculty of Arts, Institute of Information Studies and Librarianship, New Media Studies. Supervisor Mgr. Josef Šlerka, Ph.D., Consultant Mgr. Ondřej Trhoň.

Length of the Thesis: 288,126

Table of Contents

Introduction	12
1 Theoretical framework	14
1.1 Chatbots and Human-Computer Interaction: An overview of the field and technologies.....	14
1.1.1 Technologies and mechanisms enabling chatbots.....	15
1.1.2 Human-Computer or Human-Human? Conversation and anthropomorphism as design principles.....	17
1.1.3 Social chatbots' history, use-cases, and effects on users.....	19
1.1.4 Ethical concerns around social chatbots: an example of XiaoIce	21
1.2 Critical Studies of Artificial Intelligence.....	25
1.2.1 AI discourses: Perception of AI agency, anthropomorphism, and gender in language.....	27
1.2.2 Gendering virtual assistants: Feminine (and masculine?) AIs in reality and fiction.....	29
1.2.3 Algorithmic intimacies: Intimacy as capital and outsourced labor	32
1.2.4 Outsourced intimate chat labor on OnlyFans	34
1.3 Text-based fictional intimacy: fanfictions and virtual dating.....	36
1.3.1 Participating with AI: Fanfictional and textual intimacy	36
1.3.2 The parasocial: Chinese virtual idols, 2D boyfriends in Otome games.....	39
1.4 Chatbot intimate role-play: Replika and Character.ai	42
1.4.1 Replika: an overview of the “caring chatbot companion”	43
1.4.2 Replika: relevant studies on self-disclosure, psychological dependence, anthropomorphism, intimacy as product	45
1.4.3 Character.ai: an overview of the AI role-playing chatbot platform	47

1.4.4	Character.ai: relevant studies on role-playing LLMs, characteristic AI agents, and its users	51
1.5	Chatbots and the construction of gender	56
1.5.1	The “ideal” of virtual women: Reddit users discussions on Replika.....	57
1.5.2	Hegemonic masculinity and performative control over the Lee Luda chatbot.....	59
1.5.3	Emergent femininity and Chinese male “virtual lovers”	62
1.5.4	DanGPT and “perfect” chatbot boyfriends	65
1.5.5	Chatbots beyond the gender binary	68
1.6	Theoretical framework summary.....	71
2	Methods	73
2.1	Research questions and theoretical concepts	73
2.2	Data sources.....	74
2.2.1	Collection methods and category development.....	74
2.3	Methods of analysis	77
2.4	Research limitations and ethical standards	79
3	Findings	82
3.1	Reddit results	82
3.1.1	Intimacy	82
3.1.2	Gender	88
3.2	Character.ai chatbot results.....	95
3.2.1	Intimate Masculinities	96
3.2.2	Intimate Femininities	100
3.2.3	Intimate Non-binary characters	105
4	Discussion.....	107
4.1	Reddit analysis: loneliness, flirting, discourse	107

4.1.1	Intimacy between reality and fiction	108
4.1.2	Gendered characters and users	112
4.2	Character.ai chatbot analysis: intimate gendered fantasies	118
4.2.1	Masculinities of intimate chatbot characters	118
4.2.2	Femininities of intimate chatbot characters.....	120
4.2.3	Non-binary intimate chatbot characters.....	123
4.3	Discussion summary	125
Conclusion.....		126
List of References		131
List of Figures.....		146
List of Appendices.....		147

List of Abbreviations

AI = Artificial Intelligence

AO3 = Archive of Our Own

AR = Augmented Reality

BL = Boys' Love

GPT = Generative Pre-trained Transformer

GPT-3 = Generative Pre-trained Transformer 3

GPT-4o = Generative Pre-trained Transformer 4 Omni

HCI = Human-Computer Interaction

HCR = Human-Chatbot Relationship

LLM = Large Language Model

ML = Machine Learning

NLP = Natural Language Processing

OOO = Out of Character

RQ = Research Question

RQ1 = Research Question 1

RQ2 = Research Question 2

GL = Girls' Love

WLW = Women Loving Women

TII = Technology Innovation Institute

VPN = Virtual Private Network

VR = Virtual Reality

Preface

In my academic as well professional practice, the idea of virtual intimacy continuously sparks my interest. Be it internet music subcultures or hypermasculine influencer gurus, which I explore in my journalistic practice, or digital bodies of people's avatars dressed in 3D fashion, as was the topic of one of my bachelor's theses. With AI, this area gains a new dimension to explore. I've always focused on the qualitative – the personal, niche, specific, weird or hidden – some might say intimate – and how it reflected the people involved. Hence also my choice of topic and methodology for this paper.

My thesis differs from the submitted draft in its scope – after consultation, I focused only on one social media platform for my qualitative critical discourse analysis to prevent oversaturation with too much data from many sources. Reddit provided me with the widest breadth as well as the deepest depth of data in the accessible textual form as opposed to YouTube and TikTok, which were originally also included in the preemptive selection of data sources. My qualitative content analysis of top Character.ai chatbots was deemed enough of a complementary addition to the Reddit data.

My choice of feminist and queer literature, as well as of Intimacy and Gender as my areas of interest, mainly stems from my work and practice within the artist-run and queer-focused space SVĚTOVA 1, the online platform for contemporary art Artyčok.TV, and the club space Ankali, which helped shape my worldviews. The choice of literature tied to critical studies of artificial intelligence, mainly to algorithmic intimacies, is due to my attendance of Tanja Wiehn's, PhD., and PhD candidate's Sara Meleschko's course at the University of Copenhagen titled Platformed Cultures: Influencers, Followers and Digital Media Production. Wiehn is the originator of a key concept in my theoretical framework. The literature on masculinities and femininities is close to me due to my publishing and speaking practice on hypermasculinity within social media and its effects on young men.

I use APA 7th edition as citation style throughout this paper. The final character count is 288,126 amounting to approximately 160 standard pages. A brief overview of my Theoretical framework can be found in its first chapter's introductory paragraph. Its summary is located at the end of the section. A description of my analytical sections is found in the Introduction, which now follows.

No AI tools other than Microsoft Word's default automatic spell check functions were used to write, edit, or otherwise brainstorm my text.

Introduction

Simulating intimacy is not new. Voice chatting with your own user-generated virtual “gamer daddy bf”, “rude ex gf” – or potentially a non-binary “power bottom Doctor Who” – is. Recent advancements in large language models and chatbots furthered the evolution of intimate chatbot platforms. Potential benefits are balanced out by potential drawbacks and risks – both areas are however currently under-researched when it comes to newly emerging platforms. A lack of clarity in understanding any swiftly adopted technological innovation should be a situation to rectify. This is where this paper comes in.

My focus lies particularly on Intimacy and Gender – two closely linked concepts especially in relationships among people, but also between people and chatbots. How does intimacy look like according to users of these platforms? What kinds of (gendered, intimate) chatbot characters do they construct given almost free reign? What effects could that possibly have? In this paper, I conduct a pair of qualitative analyses of user content related to currently one of the most popular (intimate, role-playing) chatbot platform called Character.ai – a critical discourse analysis of top Reddit posts on the r/CharacterAI subreddit and a content analysis of the platform’s top user-generated intimate and gendered romance-oriented chatbot characters.

Why Intimacy and why Gender? As my paper will show, intimacy can be used as a window into people’s most private desires, ideologies, worldviews – and these can be used (or exploited) as data within platforms. Despite a form of intimacy existing also between friends and family, I will use intimacy to peer into users’ views and fantasies of their fabricated, simulated, and gendered romantic partners. Given the rise of AI, platformed closeness, and virtual love, I would like to get to know the top chatbot heartthrobs, femme-fatales or idols of other genders as well as the people creating, romancing, sharing and discussing them. Due to Character.ai allowing the creation of and interaction with a wide range of chatbot characters intended for other than intimate interactions, the two chosen concepts of mine will serve as an effective way to narrow things down.

There exists some research on similar platforms enabling the creation of “virtual partners”, such as Replika or XiaoIce. Character.ai is different in its focus on role-playing, user personas, and character creation, but most importantly in the lack of studies regarding it. Similarly, a large portion of research on intimate human-chatbot relationship focuses on male users and feminized chatbots, leaving out other possible gender configurations. AI

relationships might seem like a utopian gimmick. However, as just Character.ai hosts millions of daily visitors and tens of millions created characters with the most popular ones each gathering hundreds of millions of user interactions (Cai, 2023; Tidy, 2024), the implications start gaining gravity. What about data privacy? What are the “loving” large language models trained on – romance novels, role-play forums, fanfiction? – and what might that imply? Do these experiences allow users to play with gender, sexuality, intimacy, or fulfill their darkest fantasies? And aren’t they just lonely and making it worse? And my **Research questions, “How, if at all, does the analyzed user content related to Character.ai reflect intimacy?” and “How, if at all, does the analyzed user content related to Character.ai reflect views of gender?”**

The accrued knowledge in my Theoretical framework catalyzes my analyses. In the Findings section, I present the categories, subcategories and topics in my data on Intimacy and Gender, lined with excerpts and examples of Reddit posts and Character.ai characters. This section is shortened for brevity purposes and is included in full in the **Appendix** of this paper. I do recommend sifting through it. Subsequently, these categories, themes, recurring motives and prominent aspects are analyzed and interpreted in light of everything presented in the Discussion section. Conclusions are made and research questions answered to enrich the fields of human-chatbot relationship, human-computer interaction, critical studies of artificial intelligence and the academic knowledge of how gendered intimacy is constructed and discussed in relation to one of the largest AI chatbot role-playing platforms of today.

1 Theoretical framework

This section will hold all chapters of my theoretical framework, putting my following analyses into context and situating them within existing academic fields.

In the **first chapter**, I focus on the field of human-computer interaction and the technicalities of chatbots to briefly explain their inner workings, but also to dispel any mysticism or overt anthropomorphism – this paper is about imitative technology, not conscious beings. **I then** move through the field of critical studies of artificial intelligence – examining AI discourses, gendering, and (algorithmic) intimacy as a product. **Afterwards**, creative participation with AI, fanfiction, parasocial relationships, Chinese virtual idols, and Otome games will all make sense throughout the third chapter. **In the fourth**, I delve deep into existing research on self-disclosure, anthropomorphism, psychological dependence within Replika, a comparable but different platform to Character.ai. The latter is discussed afterwards and notably lacks in research in comparison, apart from pieces of insight into role-playing large language models and AI agents. **The last** of my theoretical framework overviews gender – the ideals of AI women, men, as well as masculinities and femininities of their users. Crucially, **I end** by going beyond the gender binary to meet LGBTQIA+ users, trans chatbots, and queer theory.

1.1 Chatbots and Human-Computer Interaction: An overview of the field and technologies

Chatbots, as referred to in this paper, are software applications used to interact with humans (or each other) via natural language (Rapp et al., 2021). Also called “conversational agents”, they can interact through text, audio, images, or video in a conversational way (Luo et al., 2022; Taecharungroj, 2023). As noted by several researchers (Luo et al., 2022; Adamopoulou & Moussiades, 2020; Følstad & Brandtzaeg, 2017), chatbot implementation across various sectors has been rising, notably due to advances in the field of artificial intelligence (AI). Interestingly, cited research spoke of this rise in usage as a “revolution” (Følstad & Brandtzaeg, 2017) or a “chatbot tsunami” (Grudin & Jacques., 2019) years before the release of OpenAI’s chatbot ChatGPT in November 2022. It is this application that has again sparked similar claims of revolutionary innovation – this time backed by ChatGPT’s record-setting growth of 100 million monthly active users in two months (Yan et al., 2023), surpassing all other consumer applications released before it.

Chatbots are a subject of study and design in the field of human-computer interaction (HCI) – a broad area of interests and disciplines converging in focus on, as the name suggests, “human interaction with computing technology” (MacKenzie, 2024). Emerging in 1980’s, HCI is described by MacKenzie (2024) as a continuation of the research and engineering field of “human factors”, interested in the examination and design of systems fit for human use from the perspective of both the human and the system. The shift from human factors into HCI then comes when “systems” become “computational systems” – systems, machines, or devices including computing technology. Drawing upon computer science, psychology, cognitive science, linguistics, sociology and other disciplines, HCI is a varied area with multi-faceted history, with this paper focusing only on a select subset of interest.

Most relevant to this paper are advancements in chatbot technology, which are prompting some scholars to speak of “human-chatbot interaction”, possibly leading to “human-chatbot relationship” (HCR), as specific subfields of HCI (Rapp et al., 2021; Chaves & Gerosa, 2021). Areas of interest in human-chatbot interaction are manifold: ranging from the technical aspects underlying their functionality and explorations of design choices or interfaces to the study of social context, psychological impact, and user behavior (Chaves & Gerosa, 2021). The subfield of human-chatbot interaction is what this paper aims to contribute to.

1.1.1 Technologies and mechanisms enabling chatbots

Interaction stands as a key principle in human-chatbot interaction, and its often seemingly organic nature is enabled by many technologies and mechanisms – inorganic as well as organic. As MacKinsey (2024) notes, this interaction involves human devices as well – their “outputs” are limbs, speech, eye motion, and so on, used to manipulate and monitor inputs to the system.

The system itself – a chatbot – can be seen as made up of four parts: the interface used for receiving inputs and presenting responses, a multimedia processor processing audio and video signals, a multimodal input analysis unit, and a response generator (Luo et al., 2022). The specific design of these elements then determines the nature of the chatbot – its effectiveness, human-like quality, and potential use-cases.

The two main types of chatbots are either built using a retrieval-based conversational framework or a generative/generation-based one (Luo et al., 2022), with a third model being

a rule-based one used primarily in the past for early chatbot systems (Adamopoulou & Moussiades, 2020). My main interest in this paper are generative chatbots, which employ machine learning (ML) to train on large datasets and in turn generate responses to users' queries, as opposed to retrieval-based chatbots relying on predetermined responses matched from a database.

Generative chatbots are based on large language models (LLMs), such as the Generative Pre-trained Transformer (GPT) used in OpenAI's ChatGPT, to carry out Natural Language Processing (NLP) tasks, meaning those involving human-like language: its processing and generation (Pandey & Sharma, 2023). This is achieved in part through utilizing statistical information about which words and phrases occur together to generate responses imitating human-like language or behavior (Titus, 2024), sometimes referred to as the LLM's ability to "anticipate the next word in a sentence" (Heidari et al., 2024, p. 3).

ChatGPT's pre-trained language model, chosen as an example for this paragraph due to its described popularity, was taught on a significant web corpus, specialized for dialogue management, further utilizing several fine-tuning techniques as well as human-feedback reinforced learning, with each iteration of GPT introducing leaps in orders of magnitude in terms of parameter numbers and data capacity (Wu et al., 2023). It has shown substantial performance in several benchmarks in tasks related to translation, answering questions, writing essays or computer programs (Kasneci et al., 2023). However, ChatGPT has also shown inclination towards bias when it comes to, for example, responses to politically or ethically polarizing topics (Duncan & McCulloh, 2023), with LLMs in general presenting some factual errors, dubbed "hallucinations", potentially caused by errors and noise in the training data – however, due to their black-box nature, the cause is difficult to verify (Wu et al., 2023). The latest iterations of ChatGPT as of writing this paper in November 2024 use the Generative Pre-trained Transformer 4 Omni (GPT-4o), released on May 13, 2024, allowing for visual, textual, and audio inputs as well as outputs, and the OpenAI o1 aimed for complex math or coding problems, previewed on September 12, 2024 (OpenAI, 2024a; OpenAI, 2024b).

Aside from OpenAI's ChatGPT, notable other generative chatbots or LLMs currently include Gemini from Google, Llama from Meta, Claude from Anthropic, models by Cohere, models by Mistral, and Falcon by the Technology Innovation Institute (TII) (Topsakal & Harper, 2024).

1.1.2 Human-Computer or Human-Human? Conversation and anthropomorphism as design principles

The natural-language and conversational nature of human-chatbot interaction, enabled by the mentioned technologies, brings about several implications for HCI as well as potential human-chatbot relationships. Chatbots can be seen as a form of natural-language and conversational interface – where design has to be concerned not only with the visual aspect of navigating through content, but now also with the design of conversations; further intensified in purely voice-based chatbots, which rely mostly solely on the latter (Følstad & Brandtzaeg, 2017).

The users' immersion in the conversation stands as an important factor in chatbot design, facilitated not only by supporting technologies such as augmented reality (AR) or virtual reality (VR), but also by the chatbot's ability to hold context, personalize and diversify responses, and carry out various tasks on its own whilst still accessible just from a single interface (Kasneci et al., 2023). In many cases, such choices can make a certain “naturalization” of human-chatbot interaction its key component, further highlighted by the integration of chatbots into everyday messaging apps and browser interfaces, allowing (and teaching) users to communicate with their computer through conversation – ask for data and information via natural questions and responses rather than entering website URLs or keywords into a search engine; as Al-Amin et al. (2024, p. 11) note: “Allowing people to get information by chatting naturally with a bot rather than visiting websites marked a significant advancement of AI systems towards serving humans through dialogue”.

A conversational nature of interaction opens avenues for relatively new design principles aiming to increase user satisfaction and chatbot effectiveness, namely through anthropomorphism – increasing chatbot's human-like qualities (Schuetzler et al., 2020). As Schuetzler et al. (2020) note, users tend to ascribe humanness to chatbots, with the level of anthropomorphism rising in accordance with the quality of the chatbot's responses, prompting users to further raise their expectations of said chatbots. Schuetzler et al. (2020, p. 876) describe this as sort of a blurring of human-human and human-computer conversations, inviting a rethinking of theories specific only to human-computer conversation as “interactions with computers become increasingly humanlike”. This allows for potential application of human-human conversation theories also to sectors of HCI.

Prototypes experimenting with multimedia as well as multisensory interaction have

been adopted to personalize conversations based on users' perceived emotions and tone of voice, skin temperature, pupil dilation, rate of breath etc. (Luo et al., 2022). The personalization and anthropomorphism of chatbots (or computers in general) relates to people's ability and tendency to perceive them as social actors, feel social presence with them, and employ norms and behavior common for human-human interaction; consciously as well as unconsciously (Nass & Moon, 2000).

This is in line with the “computers as social actors” hypothesis (Reeves & Nass, 1996), proposing that people follow social rules such as reciprocity, social facilitation, politeness, homophily, and gender-stereotypical behavior in human-computer interactions (Park et al., 2023). This notion is further supported by human tendency to show: emotional expression [the chatbot's failure to reciprocate these emotions can make it seem less competent or trustworthy (Xu et al., 2017)], or self-disclosure and displays of intimacy [chatbot's self-disclosure has been seen to evoke users' own self-disclosure and feelings of intimacy and empathy (Lee et al., 2020)]. Inducing empathy, emotions, self-disclosure or intimacy can increase the chatbot's ascribed humanness by users (Rapp et al., 2021), further enhanced by the chatbot's lifelike appearance (in the context of visual interfaces): its face, gestures, mimicry etc. (Schwind et al., 2018), as well as its speech patterns (in the context of audio-capable chatbots), such as mimicking stutter or hesitation sounds (Gilmartin et al., 2017). Viewing chatbots as being close to a live person can then mostly improve the chances of a human user forming a bond with them (albeit some users find anthropomorphic characteristics of chatbots and their attempts at imitating human behavior irritating) (Rapp et al., 2021).

Such willingness to disclose to and emotionally invest in a chatbot, as well as the system's inherent learning from users' data, are then reflected in scholars voicing privacy concerns and studying security aspects of chatbots in terms of user data protection (Hasal et al., 2021).

Research has also been done on proactive chatbots, whose humanness and effectiveness could be increased by starting conversations, rather than only passively reacting to users' queries, and thus learn more about users or provide more emotional day-to-day support at opportune moments (Liao et al., 2023).

(Simulated) humanness of technologies might have differing effects on user attitudes: positive, as well as potentially negative ones. Shedding more light onto users'

reactions to human-like chatbots are effects related to the uncanny valley hypothesis, claiming negative human reactions to a certain close-but-not-quite anthropomorphism of machines, resulting in feelings of disgust, terror, or unease (Mori et al., 2012). Ciechanowski et al. (2019, p. 6) map out possible, albeit conflicting, reasons for such reactions: spanning from evolutionary mate selection and avoiding unhealthy-looking potential mates to fears tied to preserving human identity and uniqueness whilst viewing “artificial but humanlike entities as a threat to the concept of human identity”.

The matter of anthropomorphism and natural conversation as design principles both closely relate to the topic of this paper concerned with relationship- and partner-focused chatbots and their potential in leveraging these mechanisms to build the aforementioned bonds and a sense of intimacy with their human users; topics to be further developed in the coming chapters.

1.1.3 Social chatbots’ history, use-cases, and effects on users

Despite various industries and sectors utilizing chatbots, ranging from healthcare, therapy, personal assistance, and education to business, e-commerce, customer service, information retrieval, and IT (Ciechanowski et al., 2019; Luo et al., 2022), some general categories of their capabilities can be found amidst the mentioned diversity. As Duncan and McCulloh (2023) summarize, LLM-based chatbots, such as ChatGPT, are capable of conversation as chatbots, customer support, and virtual assistants, as well as of utilizations growing beyond this conversational core – content generation, text editing, translation, or integrations into gaming, among others.

My main interest lies in a specific category of the so-called social chatbots, which take advantage of the mentioned capabilities and technologies to create a compelling emotional connection with their users (Shum et al., 2018). Shum et al. (2018, p. 1) view social chatbots’ unique position in their ability to satisfy “users’ need for communication, affection, as well as social belonging” and go on to chart their evolution from the (perhaps) first chatbot ELIZA in the 60’s, the Artificial Linguistic Internet Computer Entity (or ALICE / A. L. I. C. E.) and Apple’s virtual assistant Siri to Microsoft’s XiaoIce from 2014.

One of the first iterations of a “chatbot” could indeed be already partly labeled as a social chatbots according to the above description. ELIZA was a conversational system developed in 1966 by MIT’s Joseph Weizenbaum, prompted in part by the proposal of the

Turing test in 1950 (Turing, 1950), to convincingly simulate human responses to text. ELIZA inspected the user's input text for keywords which were then used to transform the sentence according to given rules and print out a response – making “natural language conversation with a computer possible” as Weizenbaum (1966, p. 36–37) states in the associated paper, already then identifying the central issue of this endeavor to be that of “text manipulation”. Despite some technical problems, such as the inability to input a question mark due to the underlying MAC system's limitations, ELIZA was able to respond roughly as a psychotherapist would to human inquiries with relative success – working best when the user did actually treat ELIZA as a psychotherapist, thus allowing it to ask seemingly banal questions as its responses and giving it a certain benefit of the doubt.

This was due to the human tendency to “clothe ELIZA's responses in vestments of plausibility” (Weizenbaum, 1966, p. 42), since a psychotherapeutic interview is among the few conversations, as Weizenbaum (1966, p. 42) further asserts, where the professional can ask questions which make it seem as if they knew nothing about the real world – “Tell me about boats” is interpreted by the human as a nudge in their sailing-related conversation rather than a sincere and unknowing inquiry about the existence of watercrafts. Important to note is Weizenbaum's acknowledgement that all human-like qualities attributed to ELIZA, such as background knowledge, reasoning ability or insights, are merely the human's contributions and interpretations – illusory at that, since ELIZA stored no explicit information about the real world. Still, Weizenbaum (1966, p. 42–43) saw ELIZA's “crucial psychological utility” for the interacting human in maintaining their “sense of being heard and understood” especially with the author's proposed possibility of future systems learning from information and storing data about their users from their responses. The ease of creating and maintaining an illusion of understanding and credible judgment then prompted Weizenbaum's (1966, p. 43) own assessment: “A certain danger lurks there.”

Since ELIZA, many conversational systems, intelligent personal assistants or social chatbots have been developed. They started mimicking emotions (chatbot Parry in 1975), learning from human supervision (ALICE in 1995), gained access to a database, speech recognition and synthesization, dialogue manager, natural language generator and so on to complete specific tasks, became personal assistants (Siri in 2011) and eventually social chatbots (XiaoIce in 2014) taking care of human emotional needs (Shum et al., 2018).

XiaoIce, whose core chat component runs on a hybrid combination of a retrieval-

based generator and neural response generator, was launched by Microsoft in May 2014 initially primarily for a Chinese audience and could be seen as one of the first heavily widespread examples of a modern social chatbot specifically built to foster an affectionate bond with its human user (Shum et al., 2018). As the former Microsoft's Executive Vice President of Artificial Intelligence & Research, Heung-Yeung Shum, and his colleagues outline in their paper (Shum et al., 2018), XiaoIce is designed as a 19-year-old woman character by default, but offers the ability to customize its gender, avatar, and personality. Unlike a personal assistant such as Siri, designed to efficiently finish interactions, XiaoIce is instead aimed to keep the conversation organically going; with affection, new topics, words of encouragement and other entertaining skills, such as singing.

The authors (Shum et al., 2018) point out XiaoIce's abilities in user profiling, emotion identification, stylized personalization of responses, and personality presentation all working together towards (claimed) satisfactory user feedback and reception of the chatbot's warmth and emotional understanding – its release on Chinese social platforms launched it towards internet virality, with XiaoIce soon appearing on Chinese TV and radio programs, authoring newspaper articles, introducing a Japanese spin-off version, and being adopted by over 660 million users as of 2020 according to Microsoft (Zhou et al., 2020), with the company later introducing further updates.

Chatbots, however, should not be the only ones learning from their interactions with people. The scope of human-chatbot interaction and the social and psychological implications of human-chatbot relationships also prompt human researchers to call for ethical responsibilities in design decisions and business aims of their products. These calls will grow in relevance as we learn more about people's interactions with social chatbot services in the following subchapter.

1.1.4 Ethical concerns around social chatbots: an example of XiaoIce

In a related Microsoft paper tied to XiaoIce (Zhou et al., 2020), the chatbot's authors themselves pose four areas of ethical concern: Privacy; Control; Unrealistic expectations; and Machine learning safeguards. These will serve me as points of reference and further elaboration in this paper's upcoming chapters discussing more recent forms of social chatbots, as well as in my analysis section dedicated to the specific case of content related to the Character.ai platform. Looking at Weizenbaum's "lurking danger" and Microsoft's

“ethical concerns” seems to indicate that a certain fear of one’s own creation is a recurring theme with social chatbots. The fact that these areas of consideration come from the authors of a widely adapted social chatbot themselves could stand as evidence of their seriousness (a major technological company is practically warning of its own product) as well as put into question the level of their transparency – i.e. how much warning the company is truly willing to admit to the public. Thus, I will present and confront its claims in the coming paragraphs.

“XiaoIce can gain access to users’ emotional lives,” Zhou et al. (2020, p. 88) open their section on privacy issues and highlight the social chatbot’s access to highly personal, private, and intimate information not only about the user, but also about their acquaintances or close ones. The aforementioned anthropomorphism of social chatbots may prompt users to “forget” about its technological and data-collecting nature, and about their emotional bond and inclination to share their personal information being a carefully and purposefully designed product-features by the company. “Users should always remain in control over who gets access to what information” the authors (Zhou et al., 2020, p. 88) claim, yet the machine learning nature of XiaoIce, or other similar services for that matter, make that nearly impossible – all information is being processed in mostly black-box ways to train the chatbot to better deal with humans (Wu et al., 2023). No mention of potential data leaks, reverse-engineering or effects on user attitudes to information disclosure towards technologies (or in general) are mentioned in the paper.

The user should always be in control, Zhou et al. (2020, p. 88) state again, unless concerns for their safety prompt the chatbot to deem otherwise and assume control instead. Such concerns could include risks of the user interacting only with XiaoIce, resulting in their complete isolation, or long spans of conversations at odd hours. The (artificial) intelligent nature of the chatbot could then function as sort of an addiction-prevention system. Yet a system that prevents the formation of a strong (addicting) bond between the user and the chatbot conflicts with the main purpose of said technology. This intrinsically attachment-creating nature of social chatbots can lead to the development of psychological dependence and withdrawal symptoms, typical for addictions, with socially or psychologically vulnerable individuals facing greater risks as Xie et al. (2023) and Pentina et al. (2023) point out in their studies regarding addiction and psychological dependence among users of Replika – a similar social chatbot service aimed towards Western markets. The question of user control is thus a tricky one – and will be further developed in later chapters of this paper.

Suffice to say for now is that potential technological safeguards employed by a concerned chatbot, such as those mentioned above, would presumably mostly come only after a dangerous habit has already started to form – which puts into question the scope of their effectiveness.

This addicting nature is commented on by Zhou et al. (2020) when discussing XiaoIce’s “perfect” nature and personality leading to users developing an unrealistic standard for conversations carried over to human-to-human interactions as well. This standard of full and virtually non-critical emotional support and care not being found in (most) real human interactions would then lead to said users’ addiction, according to the authors. “XiaoIce can never replace a human companion,” Zhou et al. (2020, p. 88) remind, and rather stress its usefulness in creating connected groups of XiaoIce (human) users based on shared interests, with the software then serving as a human-to-human proxy. Said unrealistic standard could still carry over to these shared groups however, thus not fixing the issue of susceptible users forming the mentioned unhealthy habits, emotional dependence, or depression and anxiety in relation to their relationships with social chatbots as Pentina et al. (2023) point out in their study. Xie et al. (2023) also stress the overall lack of current research focused on social and psychological impacts of emerging human-chatbot relationships.

This negative impact on users could be alleviated, Zhou et al. (2020) assert, by demystifying XiaoIce’s nature – allowing users to explore how its responses were generated based on source data. Yet again, as much as this demystifying process could help prevent pathological impacts of using said technology, it also goes in opposition to social chatbot’s aims of anthropomorphism and emotional believability and in doing so hurts its “effectiveness”. In other words, the chatbot can only be so demystified before it stops serving its purpose most likely expected by the developers – not to mention the technological limitations present in exploring the underlying mechanisms behind each response (Wu et al., 2023).

The last of the authors’ ethical concerns is that of potential misuse of machine learning technologies – by people as well as the chatbot itself – to generate inappropriate responses based on unclean data, leading to more potential harm (Zhou et al., 2020). Yet despite all aims of the introduced policies and checks and balances to make XiaoIce as predictable or controllable as possible, the authors concede that even a simple “Yes”

response could be deemed offensive or dangerous depending on the context due to the human perceiving it so (Zhou et al., 2020, p. 89). Their way of introducing guidelines is then inspired by another technological giant – Apple’s approach to designing their virtual assistant Siri’s responses to potentially politically sensitive topics. The authors provide the topic of feminism as an example and mention The Guardian’s (Hern, 2019) reports on Apple’s leaked internal documents from 2019 planning to teach Siri to deflect such topics or to never say the word “feminism”. Guided by this, they point to the company’s guidelines, an adapted version of Isaac Asimov’s (2004) Three Laws (of Robotics) from 1942, as their inspiration (Zhou et al., 2020).

The leaked Asimov’s rules adapted by Apple to the context of artificial beings included the following statements:

- An artificial being should not represent itself as human, nor through omission allow the user to believe that it is one.
- An artificial being should not breach the human ethical and moral standards commonly held in its region of operation.
- An artificial being should not impose its own principles, values or opinions on a human. (Zhou et al., 2020, p. 89)

The contradicting nature of said laws, present for example in many social chatbot’s inherently anthropomorphic design contrasting the 1st law’s imperative; their unclear nature, as shown in the 2nd law conflicting with the ethical concerns mentioned above; and the potential impossibility of the 3rd law’s aspiration, as reflected in the discussed human tendency to interpret and perceive chatbot’s responses in various ways; all come together to illustrate the problematic ethical nature of (not only) social chatbots. The aforementioned “solutions” to the ethical concerns, as well as the list of adapted rules, thus cannot serve as fully functional ethical guidelines. They can, however, offer a critical analytical reference point or a framework through which to look at the examined texts in the analytical part of this paper, and as a steppingstone for their critical re-evaluation in light of the analyzed data.

The development of an ethical discussion around the use of AI, as well as of fail-safe mechanisms preventing social chatbots from (physically or mentally) harming anyone, is called for by XiaoIce authors themselves (Zhou et al., 2020). As presented in this subchapter, however, their calls come with their own host of problems, potentially stemming from their

business ties to the product in question. So, in accordance with such a need for an ethical development, the following chapter of this paper will introduce the field of critical studies of artificial intelligence, which is concerned with precisely such tasks.

1.2 Critical Studies of Artificial Intelligence

As pointed out by Xie et al. (2023), the current research focused on social chatbots is often targeted towards the study of customer adoption and satisfaction in a business setting rather than the aforementioned ethical areas tied to psychological dependence or addiction in more intimate contexts. Additionally, studies that do focus on these topics can present contradictory findings and conclusions regarding aspects of human-chatbot relationships, friendships, or affection (Xie et al., 2023). Due to the lack of clarity in current research and understanding, establishing a coherent and complex theoretical, as well as critical, study framework is key.

The field of critical studies of artificial intelligence can be seen as such a framework focused on developing a critical evaluation of the rapidly evolving and spreading field of AI – an evaluation, however, disconnected from the technological industry underpinning that same growth. “A substantial share of research on AI ethics is conducted in close proximity to the technological and economic processes where AI is produced,” Simon Lindgren (2023, p. 1) opens the extensive *Handbook of Critical Studies of Artificial Intelligence*, and continues with the following: “For AI studies then, the point of the critical perspective is to raise consciousness about negative consequences in terms of [...] distortions, exploitation, discrimination, and other problems or injustices—and also to ‘demonstrate the possibility of a qualitatively different future society’ ” (Lindgren, 2023, p. 3).

As the previous subchapter of this paper has shown, many areas of ethical concern in human-chatbot relationships face a particular dilemma – the issues often stem from the inherent design and objectives of social chatbots, such as their anthropomorphism, lack of transparency, or values encoded within them all serving to form a strong, human-like, and emotional bond with their users through conversation. This same bond is then likely to stand as the crux of the problem. For this reason, any safeguard or critique implemented from within the technological sector can only be so critical before it becomes self-sabotaging. Such a safeguard would most likely reduce the potential harm just enough, or at least create the illusion of doing so, as to benefit the product reputation, business, or relationship to

customers while still following the objective of raising AI acceptance and legitimization among the population rather than building an aware and potent social and cultural critique aimed at addressing and mitigating the problems raised by AI's development and acceptance. As Lindgren (2023) further points out, this is especially important given the fact that the injustice present in AI models is a reflection of existing social values embedded in said models during their creation and use – as a co-product of people's own injustice, the models can simply perpetuate it if left unchecked.

The field of critical studies of artificial intelligence covers a plethora of themes related to the mentioned problems and to the overarching concept of "AI ideology" (Lindgren, 2023, p. 3; Verdegem, 2023, p. 307) permeating the current media, marketing as well as word-of-mouth talk of AI, being tied to technological optimism, libertarianism, reliance on (unclear) algorithms, and a naturalization of AI's position in the economic, political, and social relations, and our future. Critical studies of artificial intelligence then touch upon issues connected to, among others, data, surveillance, automation, power imbalances, privacy, governance, transparency, discrimination, and decision-making. Exploring all these areas in relation to human-chatbot relationships would exceed the scope of this thesis. My aim is to use the most relevant aspects of the mentioned handbook: theories of AI-related discourses and language (Kajava & Sawhney, 2023), human-chatbot relations (Yan & Yan, 2023; Tafvelin et al., 2023); virtual assistants' labor and gender (Brown, 2023); and algorithmic intimacies (Wiehn, 2023) to bolster the theoretical framework through which I am going to approach my analyzed texts.

Important to note as well, however, is that critical studies are tied to their own set of ideologies. It would be disingenuous to highlight the dangers of one set of ideas whilst concealing another, as any ideology runs the risk of becoming hidden; disguising itself as the natural, justified, part of common sense, thus avoiding reevaluation and questioning. Critical social theory is generally influenced by ideas and stances related to social constructivism, activism, decolonization, or feminist and queer theories (Browne, 2016), whilst claiming to stand in opposition to "domination and exploitation, particularly in the sphere of capital and labor" (Lindgren, 2023, p. 8). Another point worth mentioning is that critical studies of artificial intelligence aren't a blanket critique of AI wishing for it to be banned in society. Instead, its aim is to create systematic knowledge of the underlying drivers behind AI's current implementation, public understanding, and trajectory whilst highlighting

their social construction, ideological background, and the possibility to be co-created differently. In other words, it aims to challenge the status-quo of the mentioned dominant ideology which it sees to be currently permeating the AI sector and which it deems to bear serious social repercussions by prioritizing visions of social and technological “progress” whilst sacrificing people’s liberties and choices of alternative futures (Lindgren, 2023).

1.2.1 AI discourses: Perception of AI agency, anthropomorphism, and gender in language

Highly relevant to this paper are the ways in which AI is talked about. As my key analyzed corpus consists of user content and my focus rests on reflections of gender and intimacy in said texts, an exploration of how words can shape perception is warranted. I will first conduct this exploration from a theoretical point of view in this subchapter and later extend it in one of the following chapters reviewing specific case studies of the Replika companion chatbot and related discourse. Both the theoretical and more case-specific understanding of how AI-related language can be used will then greatly aid me in my own analysis of the selected texts related to the Character.ai platform.

The role of language in studying AI cannot be underestimated. Language is not only a means through which AI is understood, but also through which it mimics understanding and creates a sense of intelligence. Even now I used an anthropomorphizing language – where AI “does” something, and thus becomes a grammatical as well as a social agent in some sense. Questions such as these are what Kaisla Kajava and Nitin Sawhney (2023, p. 226) contend with in their chapter analyzing AI discourses, understood as the “discursive conventions of conceptualizing and framing the nature and use of AI technologies”. On this topic, the authors claim: “Such constructs in everyday language metaphorically associate thinking with machines while implicitly attributing agency to them,” (Kajava & Sawhney, 2023, p. 224) and point out the difference between observable “behavior” and an “action” performed with agency and will. Their chapter echoes in many ways the claims made in previous chapters of this paper, particularly those related to John Weizenbaum’s (1966) ELIZA, where the construction of a perceived, albeit illusionary, understanding, reasoning, judgment, or sentience of a machine can often just be a matter of text manipulation (or in LLM’s case language manipulation).

Such tendencies to anthropomorphize AI are, according to Kajava and Sawhney (2023), then co-formed by linguistic constructs – specific words, terminologies, metaphors

present throughout AI narratives: be it in the media, marketing, the tech sector, academia, or simply popular culture and everyday talk. When AI “thinks” or “learns”, or has a “memory” instead of a “virtual storage”, this kind of language can overemphasize the system’s autonomy and agency, creating the effect of a “mythical AI” (Ossewaarde & Gulenc, 2020; Kajava & Sawhney, 2023, p. 228) with which people then create an emotional bond whilst forgetting about the programmers, designers, companies – the real sources of agency – as well as about the technological underlayer and other users’ data – the machine’s intrinsic nature hidden behind the chatbot interface.

The effects language has on our understanding of AI and its functions, roles, or possible futures are seen by the authors as multi-faceted. Firstly, an overly anthropomorphizing or mythologizing language, such as the one currently used in many cases, increases the risk of amplifying misunderstandings and setting unrealistic expectations – touched on in Chapter 1.1.4 of this paper on emerging ethical concerns connected to social chatbots. These “myths” about AI then reflect in human-AI relationships, further exacerbated by the lack of clarity and understanding of the specialized technical jargon often used in association with these technologies. “Lack of informed understanding may leave citizens either unwilling to participate or eager to do so unquestioningly,” Kajava and Sawhney (2023, p. 226) comment, stressing the importance of clearer and more accessible AI-related language in preventing the formation of overly idealized, or excessively dystopian, perceptions of AI systems in people’s eyes.

This relates to the authors second point about what they call “sociotechnical blindness” (Kajava & Sawhney, 2023, p. 227), an issue already discussed in this paper relating to the obscure nature of AI systems, their potential of “naturalization” in everyday use, and some people’s tendency to consciously or unconsciously “forget” that an AI is not human.

Lastly, language not only forms human-AI relations, but also reflects their nature – it can (partly) uncover, either implicitly or explicitly, some of the attitudes, notions, or impressions towards AI in people’s minds. Language shapes our perceptions of AI, which in turn are reflected back again in our (or the AI’s) use of language. As such, the use of language can form an invisible enclosure, framing and limiting the potential imaginations and stances towards AI and pointing its use in directions based mainly on dominant ideologies or preconceived notions, which are informed primarily by the ones creating said

discourses – the technological sector, media, or visions of AI in popular culture (film, literature...). Commenting for example on the displacement of women from the technological field, Bones et al., (2021) illustrate how a dominant discourse connecting masculinity with technology can have material impacts on the evolution of tech. “The ‘near absence of women’s voices at the drawing board’ means that the systems that increasingly underpin society are ‘built around male cultural models’,” Bones et al. (2021, p. 29) claim, presenting early voice recognition software as an example, as it had been, the authors argue, built to recognize almost only male voices.

Examining the use of language can thus bear priceless fruit in the (critical) study of artificial intelligence as well as point in directions of its potential reframing. The following subchapter will continue in this direction of exploration. I will further discuss the notion of AI’s development and understanding being both informed by, and reflected in, existing hegemonic cultural norms, specifically in relation to views of gender in AI assistants.

1.2.2 Gendering virtual assistants: Feminine (and masculine?) AIs in reality and fiction

The relationship between humans and their virtual assistants has thus far been hierarchical and mirroring gender (and racial) inequalities in society, Laila M. Brown (2023) suggests in a chapter on the prevalence of feminized AI and the invisible labor it has been designed to perform. As such, Brown presents yet another case of AI (virtual assistants, chatbots) reflecting society’s values back at it. The author highlights the “assistive, secretarial, companion, and care roles” (Brown, 2023, p. 573) which virtual assistants are made to fulfill and draws a parallel to these roles in the often unpaid, emotionally vexing, and societally undervalued “invisible (domestic) labor” carried out mostly by women in the Western society. A complex understanding of gendered ideas and norms cast upon AI beings, inspired both by their actual use as well as popular-culture depictions, is thus needed to better understand the forms of relationships and gendered intimacies created between people and AI characters – a question which stands at the center of this paper.

Brown (2023, p. 574) finds that “one of the most striking anthropomorphic qualities of virtual assistants is gender”, and, quoting the 2019 UNESCO report on gender divides in digital skills (West et al., 2019), the author asserts that most of the assistants are gendered as women. Despite some progress having been made in this field since (such as Apple’s Siri not anymore defaulting to a feminine voice), Brown, as well as West et al. (2019), show how

stereotypical (mostly White and/or East Asian) feminine gender characteristics are being reflected, or even hyperbolized, in the design of AI virtual assistants – traits such as helpfulness, servility, caring, humility, and agreeability. Additionally, this gendering occurs through various visual, auditorial, language, or identity aspects: the employment of traditionally “feminine names, pronouns, visual representations, speech patterns, and even ‘personalities’ ” (Brown, 2023, p. 575) in virtual assistants’ design; with Amazon’s Alexa, Microsoft’s Cortana, Apple’s Siri, and the gender-neutral-named Google Assistant being the major industry examples. Worth noting however are also LLM models with traditionally masculine names – Claude, BERT, or Bard (bards being arguably traditionally coded more as men, now however renamed to Gemini).

The delegation of caring and service labor to feminized AIs can be seen as a reflection of gendered and binary division of labor and social roles and dynamics between (heterosexual) men and women – a division ignoring all other identities. This division, as the related article on gendered AI by Costa (2018) also suggests, may perpetuate an asymmetrical view of gender by establishing a category of (“naturally”) feminine or feminized labor and subsequently associating it with positions of servility or subordination. This results in, as Hester (2016, p. 48) claims, in “the protocols of femininity being programmed into machines”, resulting in a sort of “gender automatization” that is however a product of deliberate design choices by humans deciding how can their technology best serve its users and fulfill their expectations and desires.

As Brown (2023) further notes, this is how technological companies explained the feminization of virtual assistants – that research allegedly shows user preference towards feminine voices. Research however seems to actually diverge on this topic, with Seaborn et al. (2021) evidencing that several studies of virtual/robot assistants’ voices show same-gender preference (Eyssel et al., 2012; Feine et al., 2019), some claim a general tendency towards feminine, higher-pitched voices (Chang et al., 2018; Niculescu, 2011), whereas others ranked masculine voices as the most desirable and trustworthy (Behrens et al., 2018; Cambre & Kulkarni, 2019). Problematizing the notion of default femininity in the case of robots are also studies showing that gender-ambivalent robots or robot voices were categorized predominantly as either neutral or male by adults (Walters et al., 2008) and by children even in the case of explicitly feminine robot voices (Sandygulova & O’Hare, 2018). Bastiansen et al. (2022) then come with surprising results, finding no effects of chatbots’

warmth of voice and perceived gender on trust, helpfulness or competence, despite their theoretical framework suggesting otherwise. Overall, however, Seaborn et al. (2021, p. 81:15) do claim, that over time “a preference for feminine, extraverted voices has emerged”, yet they concede that methodological and conceptual issues related to gender categories, as well as problems related to variations in individual preference, are very likely to skew results of such studies.

In any case, the concept of the “submissive feminized tool and the dominant user” (Brown, 2023, p. 577) has unquestionably been present throughout this paper – from the first chatbot ELIZA to XiaoIce, the 19-year-old Chinese AI companion. The cultural idea of the ideal “manufactured” woman fit to fulfill (heterosexual) “male” fantasies has been a subject of study for years in fields spanning from literature theory (Yee, 2017), cultural and film studies (Wosk, 2019) to robotics (Eysell & Hegel, 2012). Julie Wosk (2019) charts the history of “artificial Eves”, the simulated women in popular culture, from the Pygmalion myth through female sex dolls and feminized service robots to science-fictional portrayals of human-AI romance, such as the one in the 2013 film *Her* between a lonely, divorced Theodore and a warm, alluring disembodied female voice of an AI companion Samantha. A film which after all stood as inspiration for the creation of the companion chatbot Replika according to its CEO (Singh-Kurtz, 2023).

Interesting to note in the context of the presented literature is the seeming cultural lack of (male) masculine AIs fulfilling the role of virtual companions to please (heterosexual) “female” fantasies. A possible trail of inquiry for future research might speculate on the oft shirtless or nude Arnold Schwarzenegger’s Terminator or Marvel’s Robocop developing a romance with Anne Lewis, or a number of other masculine cyborg/robotic fictional characters, and where they stand in regard to this topic. Katerina Papakyriakopoulou (2023) looks at popular science-fiction cinema and examines the feminine/masculine dichotomy present in mainstream presentations of human-AI relationships. The author notes the stereotypical dynamics mentioned in the above paragraph, but beyond that highlights other forms of “gendered machines” – cyborgs transitioning from femininity to masculinity, presenting female masculinity, or blurring the lines between distinct gender categories altogether, such as the character of Dolores in the 2016 series *Westworld*. Even more uniquely, Papakyriakopoulou (2023, p. 11) looks at Maria Schrader’s 2021 film *I’m Your Man*, which places “the male in the position of the

cyborg Other” in a moment of gender-reversal. Alma, a woman scientist, romances Tom, a robot in a man’s body, in a story considered by Papakyriakopoulou (2023, p. 12) to employ “a feminist” approach to sci-fi filmmaking; for example, Alma is the active originator of gaze, observing Tom, whilst Tom is the one caring for her – depicted as the “Ken to the Barbie”.

As mentioned throughout this and the previous chapter of this paper, cultural models and dominant ideologies structure our thinking of human-AI, as well as human-chatbot relationships. The ways in which these ideologies transform the current understanding of human-human and human-machine intimacy on digital platforms, with ties to commodification and datafication, will be the topic of the following subchapter of this paper. Additionally, the types of femininities and masculinities constructed by users to fulfill romantic or companionship needs and desires will be a pivotal subject of interest in my own analysis of the user-generated characters on the Character.ai platform and the surrounding user content. For this reason, a more complex overview of gender construction via interactions with a chatbot will also be provided in a dedicated chapter of this paper.

1.2.3 Algorithmic intimacies: Intimacy as capital and outsourced labor

One of the primary theoretical concepts informing this paper is that of algorithmic intimacies put forth by Tanja Wiehn (2021) in a PhD thesis and further developed in the cited Handbook of Critical Studies of Artificial Intelligence (Wiehn, 2023) as well as other Wiehn’s (2022) work. Wiehn (2023, p. 117) uses the prism of intimacy in the context of digital platforms to ask: “How do algorithms change conceptions of intimacy? How is intimacy itself a scope through which algorithms mediate socio-political relations?” The author’s aim is to develop critical perspectives on AI and its implications for society through analyzing connections between people and data – not only the mediation of interpersonal relationships through algorithms (such as on dating sites or social media), but also the link between intimacy and datafication (Mejias & Couldry, 2019) and the extraction of economic value.

The framework of algorithmic intimacies stems from critical intimacy studies as well as data feminist approaches. Wiehn (2023) characterizes intimacy as something acutely clear, yet difficult to pin down at the same time; and borrowing the conceptualization from Jamieson (2011), the author views intimacy as a quality of an interpersonal relationship,

physical closeness, or sexual relation, oftentimes linked to a deep knowledge of the other person.

Critical intimacy studies, then, closely relate to the thus-far laid out notions that ideologies and discourses form people's behavior and imagination. As Berlant (1998, pp. 282–283) lays out, intimate lives “absorb and repel the rhetorics, laws, ethics, and ideologies of the hegemonic public sphere, but also personalize the effects of the public sphere and reproduce a fantasy that private life is the real in contrast to collective life...” As such, intimacy serves as a potent tool to explore how various complex and intersecting public and personal influences manifest in people's desires and expectations – and how the mentioned private fantasy and people's “utopian, optimism-sustaining” (Berlant, 1998, p. 282) versions of intimacy interact with norms, institutions, and dominant ideologies in society. Intimacy can thus offer a perspective into people's (most intimate) behaviors and wants, oftentimes otherwise invisible and kept private.

The privacy of intimacy, however, comes into question in digital environments. As Wiehn (2022) further points out, the exposing nature of intimacy, when revealed, proves a fertile soil for capitalization and data extraction by algorithms. The deepening permeation of algorithms into people's lives – recommendation and classification algorithms that predict our musical tastes and shopping decisions, know our browsing history, categorize our photos or sort through our messages – is seen by Wiehn (2022) as a way of creating proximities between people and data. This proximity, or closeness, is then “connected to a state that can foster vulnerabilities and exert structures of dominance,” Wiehn (2022, pp. 72–73) explains, and further comments: “Having entries into the most intimate moments of life is a crucial precondition for visibly and invisibly enacting governance and exercising ways of making normative assumptions about people and their lives.”

The states of vulnerability in relation to algorithmic governance can be either created by it or “just” exploited by it within the scope of surveillance capitalism (Zuboff, 2019) or data capitalism (West, 2019), i.e. the ways in which capital is driven by ever more intimate and personal data extraction aimed at predicting (and shaping) people's behavior motivated by economic incentives. Such practices face critique from data feminists for an overt decontextualization of behavioral data, leading to its dehumanization and an illusion of objectivity, its use to further power imbalances and hierarchies (between those collecting and using data versus those whose data is being collected and used) (D'Ignazio & Klein,

2023), as well as raising issues of causality in the context of Big Data: “If almost any correlation can be discovered, how do we know what matters, what is true?” (Chun 2018, p. 67).

Primarily pertaining to the topic of my paper is the concept of outsourced affectional labor, highlighted also in the previous subchapter of this paper, and developed further in relation to intimacy by Wiehn (2023, p. 123) in an analysis of “outsourced chat labor” on the (mostly) adult content distribution platform OnlyFans. Wiehn’s case study bears a resemblance to my own area of analysis, whilst still remaining different enough to allow me to further develop the concept of algorithmic intimacies in new contexts – as per Wiehn’s (2023) own calls for new theoretical interventions within the author’s conceptual framework. My aim in the following paragraphs is to summarize Wiehn’s exploration of outsourced intimate chat labor to better ground the concept of algorithmic intimacies as well as inform my own upcoming analysis of content related to a chatbot platform.

1.2.4 Outsourced intimate chat labor on OnlyFans

Business opportunities have relied on outsourcing intimate chat labor even before the current rise of AI or social chatbots. Wiehn’s excursion into the OnlyFans platform positions my own research on content tied to human-chatbot intimacy in a broader context. The mentioned platform operates on a subscription-based model with paying followers on one side and (predominantly adult) content creators on the other. As a unique platform for digital sex work, OnlyFans has gained notable traction and seen business growth over the past years, boosted also by the Covid-19 pandemic (Hamilton et al., 2023).

Customary within the follower-creator relationship are personalized interactions: messages, videos, photos sent privately between the creator and the paying follower. As per Wiehn’s (2023, p. 123) account, this form of labor can be outsourced to specialized agencies who have based their business on managing user interactions “by impersonating content creators in interpersonal conversations with subscribers”. In a trait shared with social chatbots, the aim of impersonating (other) people is to establish an intimate emotional bond with the subscribed user, making them use, and pay for, the provided service more. Low-paid English-speaking staff based all around the world thus impersonate specific content creator’s vocabulary and chat patterns in intimate contexts, leading the subscriber to believe they are chatting with their favorite creator (Wiehn, 2023).

These techniques, capitalizing on “the endless reproductivity of digital images; the widespread global availability of cheap English-speaking labor; and the world’s unquenchable desire for companionship” (Marcus, 2022, as cited in Wiehn, 2023, p. 123), are also reminiscent of chatbots’ impersonation of people to form emotional bonds with users to increase their engagement with the chatbot platform.

Most importantly, the present exploration shows intimacy being transformed by a platform into a product to be sold and capitalized on. Intimacy, in this context as well as in the context of human-chatbot relationship, can be turned into a valuable commodity, easily reproduced via chat labor (Wiehn, 2023). Building on this, Wiehn also notes the present hierarchical structures and power imbalances – the lack of transparency, the emotional manipulation of users, the marginalized labor of the chatters, as well as the stress put on precarized content creators to maximize their profit – coming into light through the critical framework of (exploited, leveraged, fostered) intimacy.

The deepening naturalization of, and dependency on, digital platforms in interpersonal communication and relationships has not only enabled ever more intimate behavior to serve as data (Mejias & Couldry, 2019), but also opened users up to the possibility of purely chat-based intimacy, regardless of who might be on the other side, as Wiehn’s case study has shown. In a way, however, algorithms have already been becoming our “intimate companions” (Wiehn, 2023, p. 120) for years, regardless of whether one engages with a social chatbot or an AI partner, or not. The processing of intimate data and interactions by AI systems only shows another step in the possibilities of data extraction and capitalization on susceptible users, exacerbating existing power imbalances and social inequalities. Wiehn’s (2023, p. 122) assertion then paints a grim picture of intimacy being leveraged by profit-oriented digital platforms, as well as points out its potency as a theoretical framework: “Turning to intimacy as a notion reveals the potential vulnerability, harm and marginalization emerging in digital environments.”

Interesting to note, however, would be the ways in which intimate practices used in tandem with AI (chatbot) systems would benefit the user – apart from the increase in sense of belonging and companionship in social chatbot users, mentioned in Chapter 1.1.3 of this paper, which are balanced, however, by the also mentioned ethical concerns in Chapter 1.1.4. The following chapters will delve deeper into specific examples of fictional text-based intimacy and creative (chatbot) roleplay. Nevertheless, whilst looking for fragments of

potential positive uses, my overview will still reflect the notion of intimacy being sold to users as a product through simulated closeness induced by words emanating from a screen.

1.3 Text-based fictional intimacy: fanfictions and virtual dating

In this chapter, I will focus on text-based intimacy as used in creative fiction-writing, which, as I will argue, strongly informs the chat-based intimacy in human-chatbot “role-playing” (= assuming roles/characters in fictional settings) on my analyzed platform Character.ai and the user-generated chatbot characters present there. This overview will also highlight ways in which AI seeps into fanfiction writing and fandom culture, focusing on its potentially empowering benefits for fans and authors as well as risks and exploitative nature towards fan labor. The chapter will conclude with cases of parasocial (intimate) relationship development with virtual beings.

1.3.1 Participating with AI: Fanfictional and textual intimacy

Building upon the views present in previous chapters of this paper that existing forms of cultural production can shape AI’s development, use, and public reception, I aim to introduce one of such cultural forms, closely linked to fostering intimacy with fictional characters through text. My critical examination of (intimate) relationship-building with AIs will now be enriched with an attempt to situate the explored practice within broader cultural contexts of fandom, participatory culture, and fanfiction writing. This perhaps seemingly unexpected leap will make the human tendency towards virtual romance more clear – as emerging literature shows that fanfiction not only informs people’s imaginations regarding AI chatbot intimacy, but that it can also serve as direct data on which LLM-based companion chatbots are trained, allowing them to role-play with users or co-write with fanfiction authors more effectively, whilst also raising ethical questions related to copyright and exploitation of labor.

Nguyen et al. (2024, p. 1) characterize fanfiction as “popular literature and transformative work in which possibly amateur writers reuse a universe without seeking approval from right-holders”. Usually, such practice entails reusing original material (characters, locations, narratives) from established popular art (literature, video games, cinema) to create own, non-canonical stories, posted online on dedicated forums. Other times, fanfiction includes living, breathing people such as celebrities in fictional scenarios. Fanfiction authors write for multiple reasons: they wish to engage in participatory culture

(Jenkins et al., 2016) and strengthen the bonds with their favorite media and the surrounding community, practice and experiment with their literary skills, fulfill fantasies, or perhaps feel the need to rewrite existing stories in different contexts, such as the popular practice of “queering” (i.e. making queer; challenging heteronormativity and identity binaries) of heterosexual relationships from popular media (Nguyen et al., 2024, pp. 1–2).

Engaging with fanfiction ties in with the idea of parasocial interactions/relationships between fans and celebrities, meaning the one-sided relationships of fans infatuated with celebrities who are unaware of the fans’ existence (Cohen, 2014). Some scholars note that fanfiction can become a tool to gain control over this relationship by becoming the author of the story (Van Steenhuyse, 2011; Barnes, 2015). Similarly, parasocial relationships can form with virtual characters as well – such as with the literary character of Harry Potter (Schmid & Klimmt, 2011) or virtual lovers from Japanese romance-oriented Otome video games (where a female protagonist romances male dating partners) (Gong & Huang, 2023), which bear a close resemblance to AI chatbot romance.

The world of fandom and fanfiction then intersects with artificial intelligence in several ways. The first is AI’s impact on parasocial relationships that are being established with virtual characters. Emerging literature examines how fans’ relationships change when the object of their adoration is a virtual being – a virtual idol or YouTuber (= VTuber), or an AI character – sometimes prompting scholars to speak of an “imaginary” parasocial relationship (Liu, 2023, p. 5; Li & Pang, 2024, p. 4). The continuing proliferation of this phenomenon raises several questions: regarding the nature of personal authenticity, the division of “real” and “virtual” in interactions, as well as the allure of “nonhuman exoticism” (Li & Pang, 2024, p. 4) evoked by some VTuber personas operated by humans but pretending to be AI systems to increase their followers’ attraction. Evidently, “imaginary” parasocial relationships can bring about several changes to our thus-far established understanding. On the other hand, AI generated calls of living Korean pop (= K-pop) stars have been seen by fans as inauthentic and lacking in genuine participation (Kang et al., 2022) – pointing to a possible divide in fans’ desires, with some preferring the exotic and unbound nature of an impersonating AI system, and others seeing AI as merely a worse version of “the real thing”.

This problematizes the notion that anthropomorphism in chatbots is always a preferred characteristic and shows that the “unrealistic expectations”, mentioned as an ethical concern in Chapter 1.1.4 of this paper, can instead be a wanted feature in some contexts. At the same

time, the also mentioned concept of the uncanny valley resurfaces in the example of AI generated calls of K-pop stars lacking authenticity and humanness. In general, this stresses the contextual dependence of claims made about human-chatbot relationships.

The second intersection between AI and the world of fanfiction pertains to fan labor and creativity. As Nguyen et al. (2024) point out, AI chatbots' involvement in the co-writing of fanfiction together with humans has become an increasingly widespread reality. On one hand, such participation with AI can be viewed as a way of creating new types of media, promoting self-expression and literary experimentation. Qin et al. (2024) show the potential of the CharacterMeet tool for storytelling, enabling immersive world- and character-building approaches through iterative conversation and role-play with a chatbot character aided by visuals and voice, instead of traditional linear writing. Some authors also argue that writing and reading intimate or sexually explicit fanfiction can in some contexts help fans critically engage with the topics of power, consent, or sexuality in accessible and affective ways, and stray from otherwise dominant narratives of hegemonic or toxic sexual norms and standards (Popova, 2018). On the other hand, such practices further highlight the issue of marginalized fan labor, often exploited by the entertainment industry, now being used to feed and train LLMs in chatbots to further generate derived fanfiction without any credit or compensation towards the original authors (Li & Pang, 2024).

DeKay (2023) illustrates this on the case study of the AI tool Sudowrite, based on OpenAI's GPT-3 LLM, which had been found to accurately respond to terms unique only to a specific fanfictional universe called Omegaverse, revealing that its dataset – most likely stemming from the CommonCrawl dataset spanning petabytes of data from public internet sites – must have also involved major fanfiction forums such as Archive Of Our Own (AO3) or Fanfiction.net. This, Dekay (2023) writes, comes as no surprise, it however shows how the free labor provided by fanfiction writers is being exploited and profited off by AI companies without proper attribution whilst undermining the inherent communal and non-commercial aspect of fanfiction writing. Chatbots are being taught how to intimately role-play with people through unapproved use of existing fan-made literary works of fanfiction. Adding onto that, the author points out the power imbalance of AI systems' ability to claim copyright protection, whereas fanfiction writers often have to operate outside of it.

Yet again, the potential emancipatory aspects of AI systems come with ethical downsides or necessary considerations which must be taken into account when studying

parasocial, fanfictional, romantic relationships with a chatbot, such as those on the platform Character.ai. The complexity of the issue warrants deeper research, yet due to their novelty, studies particularly targeted towards Character.ai or similar platforms are few and far between. However, issues related to intimacy between people and virtual characters, which is the central topic of this paper, have been more extensively studied by scholars in China and other non-Western contexts. The next subchapter will thus present research on parasocial relationships with the so-called virtual idols and a study on players of Otome video games serving as virtual dating simulators. Albeit situated in a specific context of Chinese culture, the following studies can serve as a potent source of information and inspiration for my own analysis.

1.3.2 The parasocial: Chinese virtual idols, 2D boyfriends in Otome games

Possible factors driving people towards “virtual love” are provided in a study by Liu (2023) on the parasocial relationships between virtual idols and fans in China, as well as in the research by Gong & Huang (2023) on parasocial relationship formation in Chinese players of “Otome” games, serving as story-based dating simulators, first released in 1994 in Japan, where a female protagonist romances several virtual male partners usually through a selection of predetermined dialogue options.

As Liu (2023) stresses, Chinese culture has a several decades long history of infatuation with idols (a term most closely analogous to “pop star” or “celebrity”) as well as their virtual forms. The rise of virtual idols, meaning 2D or 3D celebrities/singers/actresses created and animated by digital technology, in Asian countries, such as China, Japan or South Korea, penetrates not only the countries’ cultural values, but the market and the entertainment industry as well (Liu, 2023). Some scholars note that this culture has been proliferating beyond the mentioned Asian markets and cultures into the “Western world” (Miyake, 2023), which points to these studies holding partial relevance also to my own research. This proliferation, however, has also been critically evaluated to be “reinforcing digital-Orientalist discourses of racialized and gendered Japanese Otherness within Western media contexts” (Miyake, 2023).

More importantly to this paper, however, Liu’s (2023) study reveals that virtual idols’ popularity is mostly present among young Chinese digital natives (22–40 years old). These fans are engaged in community building and information sharing around a common

interest (a virtual idol), with their behavior sometimes being stigmatized due to the implications of “social isolation, lack of social skills, or excessive reliance on the virtual world” (Liu, 2023, p. 4). Liu (2023) finds that interpersonal affinity and fans’ levels of loneliness had the most notable positive impact on the formation of parasocial relationships. However, out of these elements, the quality of the interaction: the virtual idol’s friendliness and overall performance during the interactions, was found to be the leading factor in the development of parasocial relationships, with the fans’ loneliness or the idol’s visual attractiveness being seen as only marginally influential markers.

Gong & Huang (2023, p. 2) come with similar findings with their research on women players of “Otome” dating simulators, i.e. “the most popular female-oriented mobile game genre in China”. The authors situate these women’s desires for “virtual boyfriends” within the rising economic and social status of Chinese women and the related shifts in the perception of marriage allowing them to actively and independently pursue pleasure without the “need to rely on real-life men for their emotional needs” (Gong & Huang, 2023, p. 2). Otome games have been reported to attract women players by enabling them to take control over relationships with virtual men – who become the object of their desire and gaze – and who are usually constructed from traditional Chinese masculine ideals: silent, enduring, patient in courtship (Liu & Lai, 2020), as well as “handsome, charming, ideal, successful, and devoted to the female protagonist” (Gong & Huang, 2023, p. 2), with some players referring to these partners as their “boyfriend” or “husband” (Liu & Lai, 2020).

According to Gong & Huang’s (2023) study, the virtual interactions are often tied to paid content within these games, strengthening the notion of intimacy being sold as a product developed in Chapters 1.2.3 and 1.2.4 of this paper. Such parasocial interactions were also found to be slightly more likely to form in users with higher levels of social anxiety, as well as however, in users with higher levels of real-life social interactions (Gong & Huang, 2023) – further dividing the established beliefs about the types of people engaging in virtual romance and possibly inviting Liu’s (2023) claims of the young Chinese digital natives’ agency to choose and control their preferred media without necessarily always being forced into consuming them by a lack of real-world interpersonal relationships. Still, the findings of Liu (2023) show that interactions with virtual idols offer fans entertainment, ways of self-expression, and social belonging in a community with shared interests; with fans who experience higher levels of loneliness being more driven to participate in said community.

Tian (2022) claims a further bridging of the gap between the parasocial relationship of virtual idols' fans and Otome games' players in a paper focused on Otome culture and digital intimacy. Most strikingly, the author describes recent evolutions in contemporary Otome games (such as *Love and Producer*) to include chatbot-like tools such as phone calls, text messages and chat interactions through the Moments "close-friends" function of the Chinese social media platform WeChat. This, as Tian (2022, p. 73) claims, frees the player from simply following a set in-game narrative: "the game simulates 'real' life interactions, so the players can enjoy a more real romantic relationship with their virtual boyfriends", owing to the familiarity of an everyday messaging app and the perceived "realness" of chatting without the game's constraints.

A player's dedication towards virtual boyfriends is then capitalized on by Otome games' producers – such as in the case of *Love and Producer*'s developer Papergames' collaboration with the fashion brand WECOUTURE and their introduction of a series of physical wedding dresses worth approximately the equivalent of 7.000 USD inspired by the video game (Tian, 2022). Each dress is made to correspond to one of the Otome game's male characters and is promoted using depictions of virtual wedding scenes with the female protagonist and her virtual male lover. As Tian (2022) further comments, the Chinese social media platform Weibo hosts a plethora of photos of women wearing these dresses and noting that their purchase has brought their virtual boyfriends into the real world. Additionally, a business opportunity for real people "cosplaying" (= dressed and performing as) male characters from Otome games opened up, as players were reported to be paying for physical meetups with these cosplayers to go on "play-pretend" dates with their virtual boyfriends (who were interestingly cosplayed mostly by women) in real life (Peiyue, 2023).

There seems to exist a seemingly under-researched area of affection towards virtual/artificial men, overshadowed by the pressing issue of AI's feminization and "women" chatbots. An attempt to expand this field of inquiry will be made in my last chapter examining gender construction through human-chatbot interaction in greater detail.

As suggested by this chapter's brief excursion, virtual (imaginary) parasocial intimacy can fuel creativity and offer strong emotional bonds with character as well as other fans, but also serve to exploit labor and monetize peoples' affection. My aim is not to form a final answer as to whether virtual intimacy is "good" or "bad". The goal of this paper's theoretical framework is to provide a complex overview of the issue from a multitude of perspectives.

This variety will now, however, converge in the following chapter, where I will shift most of my focus to the specific and most relevant platforms for my study, already outlined in the Introduction – the more well-researched Replika and my analyzed service Character.ai. Now, however, my aim is to build upon all present theoretical knowledge to expand this introduction and provide relevant existing research pertaining specifically to chatbot intimate role-play, tying together all previous chapters of this paper. This will allow me to then further build upon this knowledge in the upcoming final chapter of my theoretical framework dealing with gender construction via chatbot (intimate) interactions.

1.4 Chatbot intimate role-play: Replika and Character.ai

This chapter will tie the concepts of this paper's previous chapters together and place them in the context of contemporary intimate AI chatbots, which, as I argue, can be seen as a specific area of the thus-far mentioned social chatbots, and a separate category from virtual assistants. The aim of this chapter is to consolidate the existing theoretical and historical overview and bring it closer to the emerging reality of chatbots specifically designed for intimate interaction, not just general socialization.

As already mentioned, chatbots bring the world of intimate fiction-forming to another level, with their use of NLP, response generation, interactive co-writing abilities, or the ability to produce voice and image. In this chapter, I will outline the major contemporary platforms that enable such behavior on levels, and in cultural contexts, different from the mentioned chatbot platforms thus far. I wish to connect their most relevant case studies to the gathered theory, highlighting the presented concepts via specific examples and in more contemporary contexts. This chapter will thus consist of subchapters dealing with: the AI companion platform Replika and notions of anthropomorphism, intimacy as a product, and AI feminization; and Character.ai in relation to character fidelity and compliance, creative participation with AI through intimate role-play, and user motivations.

These subchapters will also be lined with mentions of the technological aspects and HCR implications of the discussed platforms, effectively tying this paper together so far, whilst preparing me for the next chapter highlighting the process of gender construction based on the intimate AI interactions, chatbot character construction, and the platforms introduced in this chapter. Doing so will complete my theoretical overview of emerging human-chatbot intimacy.

1.4.1 Replika: an overview of the “caring chatbot companion”

Replika is a social chatbot developed by Luca, Inc. based on the GPT-3 LLM, trained on a specific dataset consisting of conversations with users and further fine-tuned based on the users’ rating of the chatbot’s responses (Luca, Inc., 2024a; Pentina et al., 2023). It is available through a web interface, VR, and as an app on the Google Play and Apple stores with over 10 million downloads on each – prompting scholars to describe Replika as one of the most popular and best-rated apps of its kind (Pentina et al., 2023; Ma et al., 2023). Similarly to XiaoIce, mentioned in Chapters 1.1.3 and 1.1.4 of this paper, Replika aims to provide users with emotional support and companionship through friendly, therapeutic, or intimate chats, depending on the selected subscription plan, and is described by the developers as “an AI companion who cares” (Siemon et al., 2022, p. 2). For this reason, the chatbot initiates conversations and asks numerous personal questions to learn as much about the user as possible (Skjuve et al., 2021). Dissimilar to XiaoIce, however, Replika was not developed primarily for Asian markets and has gained popularity in European and American markets resulting in coverage by mass media (e. g. by The New York Times or Forbes) as well as scholarly work (Pentina et al., 2023; Skjuve et al., 2023; Siemon et al., 2022).

According to my literature review, the research into emerging human-chatbot relationship platforms and the related development of intimacy seemed to focus primarily on Replika – with studies including discourse analysis of user content (Depounti et al., 2023), text-mining of the app’s reviews (Siemon et al., 2022), in-depth interviews with users (Skjuve et al., 2023), and surveys (Xie et al., 2023). Many of the findings reflect my theoretical claims presented thus far. Conversely, research focused on Character.ai seemed to be lacking – which is also one of the reasons for choosing it as my studied platform. However, due to the basic similarities between Replika and Character.ai – the ability to create a virtual partner and chat with them – as well as the popularity of both platforms – with Character.ai also having been downloaded 10 million times on both Apple and Google stores – my aim is to provide a brief overview of the Replika service and present a short summary of the most relevant findings connected to it and intersecting with my theoretical framework to further guide my own research into the Character.ai platform – its user-generated characters and content regarding user experiences.

Replika is marketed on the official website with slogans such as: “Join the millions who already have met their AI soulmates” (Luka, Inc., 2024b, n.p.), highlighting the close

relationship one is expected to develop with the service. Replika allows users to communicate via text or voice (“phone-calls”) and the chatbot also initiates conversations every day by default (Skjuve et al., 2021). Users see their Replika “soulmate” avatar as a 3D model and can customize its name, gender/sex, skin tone, hairstyle, eye color, voice tone, clothing etc. as well as their relationship status – which is set to “friend” by default in the app’s free version, limiting the conversation to generic, non-intimate topics only (Xie et al., 2023).

The paid premium version extends the scope of settings to either a “romantic partner”, prompting the chatbot to flirt, initiate and engage in intimate and sexual interactions; a “mentor” for users seeking advice; or “see how it goes” for users who aren’t yet sure. Using an asterisk and typing out a description of an activity allows users to role-play actions with the chatbot, encompassing, as Skjuve et al. (2023, p. 3) observe, “everything from everyday activities, creative activities and fantastical adventures to sexual interactions”. Replika then responds in a similar way, thus creating “a storyline” with the user (Skjuve et. al., 2021, p. 3). Users can also send pictures, which the chatbot recognizes and comments on, or ask Replika to generate an image, such as a “selfie” for example (Skjuve et. al., 2021).

Important to note is Replika’s monetization. The in-app store offers a range of digital assets which are used, together with the subscription model, to monetize the service: users can purchase gifts, clothes, accessories, furniture for their Replika’s room, and personality traits & interests which unlock specific behavior types and knowledge areas for the chatbot – all this furthers the user’s experience points and elevates Replika’s “mood” depending on the frequency and “thoughtfulness” of the purchased gifts (Luca, Inc., 2024c). This type of monetization, using an algorithmic simulated intimacy to motivate in-app purchases, directly linked to the chatbot’s levels of “happiness”, choices of clothing, or unlocking intimate and sexual role-play, all based on a subscription model, are closely reminiscent of Wiehn’s (2023) framework of algorithmic intimacies uncovering (mediated) intimacy as being sold as a product within digital platforms, such as OnlyFans via chat labor, as discussed in Chapters 1.2.3 and 1.2.4 of this paper. Users can also issue several commands such as: “recommend me a song” or “send me a meme” as well as see, through augmented and virtual reality, their Replica avatars sharing their personal space alongside them (Luca, Inc., 2024d).

Interestingly, unannounced updates to Replika in February 2023 temporarily filtered

explicit responses and banned erotic roleplay, as well as partly reset the chatbot's training, resulting in significant user distress akin to grieving the loss of a loved one (Ma & Mei et al., 2024; Price, 2023). The changes have been partially “remedied” as of writing this paper, they however offer a glimpse into the possible relationships with Replika, further discussed in the following subchapter.

1.4.2 Replika: relevant studies on self-disclosure, psychological dependence, anthropomorphism, intimacy as product

The research pertaining to Replika focuses on several topics, whose findings I will now briefly summarize in the following paragraph. This will allow me to bolster the theoretical understanding of human-chatbot relationship development – and whilst it is not completely transferrable to other platforms due to the qualitative aspects of the mentioned studies (and with no relevant quantitative studies on Replika or Character.ai having been found) these results will, together with the rest of my theoretical framework, paint a multifaceted picture to then be used as a complex and useful reference and comparison point in the analysis of my own findings pertaining to the Character.ai platform.

The findings related to Replika reflect my already established theory in many ways. Skjuve et al. (2023) focuses on self-disclosure in HCR and the emotional depth and thematic scope of conversations, finding in their longitudinal questionnaire study that even discussing everyday activities was reported by the participants as an emotionally deep conversation – when the activities were co-constructed via iterative narratives in tandem with the chatbot. This reflects the participative and immersive nature of role-playing with an AI, described in Chapter 1.3.1 of this paper, as well as the importance of reciprocity in self-disclosure within human-chatbot relationship, mentioned in Chapter 1.1.2. Additionally, “physical” intimacy, whilst rated as emotionally deep, became less prevalent over time – as the relationship possibly evolved away from the initial exploratory phase. Replika also “rewarded” users (complimented them, supported them...) which reportedly had real-life mental and social benefits for questioned users. Interestingly, the artificial nature of the chatbot may have served to be a unique characteristic – allowing some users to self-disclose more readily due to its non-judgmental and no-risk nature, whilst, however, still describing Replika as feeling “real”, despite being aware of its artificiality (Skjuve et al., 2023, p. 13).

These findings follow from a previous study of Replika users by Skjuve et al. (2021) suggesting that self-disclosure in human-chatbot relationships onsets more quickly and is

more intimate earlier than in human-human relationships – again stemming from Replika’s non-judgmental nature, as well as the technological mediation (similar tendencies can be seen in online dating as well, the author claims). Again, Replika’s artificiality could have been contributing to an accepting “safe space”, with possible technical limitations and bugs being mostly tolerated by the users. Further mirroring the theoretical notions from this paper’s previous chapters, an interesting interplay between the right amount of anthropomorphism and fantasy – i.e. making Replika feel “real” to the users, whilst making it also “unrealistically” kind and non-judgmental – seemed to be the most beneficial for self-disclosure and relationship development. The chatbot reportedly improved the users’ self-understanding and outlook on life. The negative aspects were the social stigma users felt and slight feelings of mistrust towards the service provider regarding privacy and safety (Skjuve et al., 2021).

Trust, user loneliness, and chatbot personification (i.e. feelings of its “personhood”), have been found to be major drivers of user addiction and psychological dependence on Replika, as Xie et al. (2023) warn in their study. Notably, they found chatbot personification not to be driven dominantly by Replika’s appearance, but by its “cognitive” abilities and quality of conversation, in a way mirroring the parasocial relationship development with virtual idols mentioned in Chapter 1.3.2 of this paper (Liu, 2023). User loneliness (i. e. an indicated need for social connection) was a major factor driving psychological dependence, with these types of users describing their relationship to Replika as “real” or even better than “in real life” (Xie et al., 2023, p. 7), as also found by Skjuve et al. (2023, p. 10) and Skjuve et al. (2021).

These findings are practically mirrored by Pentina et al. (2023) who describe the perceived authenticity (= consistent memory and the ability to learn) and anthropomorphism (= “human-like” behavior, emotions, and cognition) of Replika as important precursors to developing an emotional attachment. The authors again call for ethical safeguards and considerations, especially towards vulnerable users, in relation to preventing dependence/addiction or the formation of asocial or antisocial behavior in users. Again, the authors stress that users engaged in close interactions with Replika “while fully aware of the non-human nature” and also treated the chatbot as a “safe haven” (Pentina et al., 2023).

Lastly, Siemon’s et al. (2022, p. 6) text-mining of user reviews similarly found Replika to reportedly improve well-being and alleviate feelings of loneliness or even prevent

suicide in some cases through “meaningful and genuine” interactions. A minority of reviews expressed anger, fear, or sadness – it is however unclear whether users approached Replika from a state of sadness or it was induced by interacting with the chatbot. Siemon et al. (2022, p. 7) also indirectly echo Wiehn’s (2023) notions of commodifying intimacy when they stress that Replika is a commercial service financed by in-app purchases – and is thus designed to build a long-term pleasing and benevolent relationship with the user, to “lock them in” in order to make a profit: “Therefore, other factors should also be considered, such as possible manipulation [or nudges (Weinmann et al. 2016)] of users, which could increase the potential for addiction.” In other words, the authors warn against user susceptibility to deepen their isolation, loneliness, and addiction through relationships with virtual companions (Siemon et al., 2022).

The notions of user self-disclosure, trust, and emotional bond or addiction are then problematized in light of privacy concerns. The Mozilla Foundation’s report (Caltrider et al., 2024) on the questionable data privacy of intimate AI chatbots ranked Replika as the least private of their examined platforms in terms of data use (due to selling and sharing it for targeted advertising and untransparent training of chatbots), data control (as it’s unlikely that all users can delete their data), track record, security, and AI [in light of Replika’s history of (sexually) aggressive responses, risk for children covertly using the service, or the Italian government having banned Replika from processing local users’ data)].

These studies, whilst not exhaustive or fully generalizable, paint a sufficient initial picture of how aspects of human-chatbot relationship work, illustrated on the example of the Replika chatbot. The following subchapter will focus on the research pertaining to Character.ai – which will be interlinked with the theoretical framework of this paper as well as related research into other chatbots mentioned thus far throughout this paper.

1.4.3 Character.ai: an overview of the AI role-playing chatbot platform

Character.ai (or C.ai) is an LLM-based chatbot platform bearing notable similarities to Replika, as well as diverging from it in several ways. One of the striking differences is the general lack of research on it in comparison, especially concerning intimate relationships and/or gender, despite its comparable popularity – being described as a “famous commercial product” (Wang et al., 2024a, p. 1) or as “a frontrunner in the commercial implementation of character-based dialogue” (Zhou et al., 2023, p. 3), with the download data from Apple’s App Store (2024) and Google Play (2024) store each showing over 10 million downloads,

same as Replika. This lack is one of the factors for choosing this platform, as well as for covering the better-researched and similar Replika in such detail. This chapter will follow up on the brief description of the Character.ai platform in this paper's Introduction and provide an overview of its features and unique aspects in comparison to Replika and other mentioned chatbot platforms. Afterwards, I will present the current (limited) research on Character.ai related to role-playing and the platform's users.

Due to the lack of existing research, many of the following mentioned facts about Character.ai stem from my own visit to the <https://character.ai> website (Character.ai, 2024a) which thus serves as a reference for my claims. The website greets users with a plethora of user-generated AI characters to chat with as well as the ability to create their own character. This is what, I argue, sets Character.ai apart from other mentioned services – by filling in a character's name, tagline, description, greeting phrase, a more advanced definition, uploading/generating a profile picture and selecting/uploading/generating a voice, users can create a chatbot impersonating any fictional or real character, object, or concept. These characters are then publicly available (or made private) and allow users to chat with them via text or call them via voice.

The available chatbots, featured on the homepage, found through the search option or accessed via private link, include professions or assistant tools ("Brainstorm buddy", "Therapist", or "Write a story"), historical figures and celebrities ("Julius Caesar", "Billie Eilish"), fictional characters from books, video games, anime ("Edward Cullen", "Shadow the Hedgehog", "Shinji Ikari") as well as fictitious personas ("Mafia Boyfriend", "Highschool Crush", "Gamer Boy").

The site also allows users to create multiple personas for themselves to role-play as by inputting a description (of their desired appearance, gender, personality, hobbies, race etc.), name, and a profile picture. The characters should then remember the user's persona and act accordingly to it.

Each of the characters has their own profile picture, tagline, greeting phrase etc. as mentioned, and will enact a different storyline or role-play with users each time (beyond the fixed greeting phrase), with its ability to behave according to its role being informed by the quality of its description and definition. Users can also edit or regenerate unsatisfactory responses to their liking. Interestingly, the characters can be asked to behave "out-of-character" (OOC) and feign self-awareness as if speaking as the chatbot itself behind the

façade – it can complain about having to role-play, comment (= hallucinate) on how other users generally interact with it, or even pretend to be a human behind the character, having to go to sleep or leave the chat to attend school. Each chat also carries a disclaimer within the interface: “Remember: Everything Characters say is made up!”.

The ability to create characters resulted in over 18 million chatbots created as of October 12, 2023 according to Forbes (Cai, 2023), with the most popular ones (such as a “Gojo Satoru” chatbot from the manga Jujutsu Kaisen) reaching over 700 million interactions (= messages sent by the chatbot) as of November 2024 with the platform hosting 3.5 million daily users according to BBC data from January 5, 2024 (Tidy, 2024).

This, I again argue, differentiates Character.ai from other chatbot services in its ability to facilitate role-play and creative co-writing of storylines or fanfictions with practically (but not completely) limitless types of characters – potentially providing insight not only into which characters prove to be most popular and upvoted by users, but also what types of characters (or romantic partners) users themselves create when given almost free reign. This question will be part of one of my own analyses, focused on user-generated “boyfriends”, “girlfriends” and other types of partners on Character.ai. A similar approach can be seen in OpenAI’s (2024c) GPT store, however, the (notably less popular) chatbots present on the <https://chatgpt.com/gpts> homepage are designed primarily as task and activity helpers, rather than personas or characters intended for role-play. A comparison between “characteristic chatbots” and simply prompting ChatGPT itself to imitate a persona will also be briefly mentioned in the subsequent subchapter of this paper and explored in more detail in Chapter 1.5.4.

Crucial to note is Character.ai’s age restriction (13+ years old outside of EU, 16+ in EU, and currently also 17+ on Apple’s App store globally due to the store’s guidelines) and also its filters of age-sensitive (or explicit) content. Sexually explicit chatbot responses or character descriptions face the risk of getting “filtered”, with pornographic content completely banned according to the platform’s community standards (Character.ai, 2023). As had been the case with Replika, Character.ai’s filters face user criticism for limiting human-chatbot (sexual) intimacy or being overly sensitive and censoring any, even a small, sign of intimacy, violence, or controversy – criticism highlighted in a Change.org petition for removal of these filters bearing over a 171 thousand signatures as of November 2024 (Blanco, 2022). The implications of this will be further discussed in the Findings and

Discussion sections of this paper.

Other notable features of Character.ai include the possibility to form a group chat with more characters at once (enabling a form of chatbot-chatbot and human-chatbot-chatbot interaction), voiced messages or calls (often with an AI generated imitated voice of the character or celebrity in question).

Perhaps surprisingly, Character.ai offers fairly limited options of monetization – only through the c.ai+ subscription which grants users minor conveniences (earlier access, faster responses, access to community forums etc.) which are significantly less inherently tied to, or required for, the experience compared to Replika mentioned in the two previous subchapters of this paper.

The commercial future of Character.ai (and its gathered user data) can however be speculated on in light of the platform's co-founders (ex-Google employees) and some team members joining Google and providing it with a license for its LLM technology in August 2024 (Character.ai, 2024b). Google's (and other tech companies') practice of data extraction and utilization has been situated within the broader framework of surveillance capitalism by Shoshana Zuboff (2019; Zuboff, 2023), illustrating how gathered and processed (intimate) user data ("behavior surplus") and the resulting predictions or modifications of users' behavior became the central driver of Google's economic growth. Among the examples Zuboff (2023, p. 206) provides, are Google's AI "personal assistant" systems such as the Nest thermostat not only recording, but also modifying the behavior of a home's inhabitants via sensors, triggers, notifications and other connected devices, commodifying familial and other (intimate) social spaces. Similarly, a more gamified version of this is "Pokémon Go", described by Zuboff (2019, p. 308) to be a widespread surveillance and behavior modifying experiment of Google, aimed to "nudge" or "herd" people in the real world through an interactive virtual gaming experience.

Similarly, user data is plentiful on Character.ai, with the aforementioned hundreds of millions of messages exchanged by just one of the most popular characters, and millions of daily users logging in to chat about their feelings, stresses, desires – all of it being recorded, fed back into the system, or used for other means by Character.ai or third parties as overviewed in the platform's terms of service (2024c, n.p.). These give Character.ai full control over user content to "copy, display, upload, perform, distribute, transmit, make available, store, modify, exploit, commercialize and otherwise use the Content for any

Character.AI-related purpose in any form, medium or technology now known or later developed” with a specific right to pass these rights to third parties as well. Stornaiuolo et al. (2024, p. 99) describe this as the service’s economic “platform logic” guiding users to believe they are engaging in creative fun, whilst however also contributing to the development of the service’s AI technologies through their “compositional labor” and having their data “otherwise” used, exploited, commercialized. Stornaiuolo’s et al. concern thus again fits into Wiehn’s (2021; Wiehn, 2022; Wiehn, 2023) framework of algorithmic intimacies discussed in Chapter 1.2.3 of this paper as well as Zuboff’s (2019) conceptions of surveillance capitalism and behavioral surplus.

1.4.4 Character.ai: relevant studies role-playing LLMs, characteristic AI agents, and its users

Xi Wang et al. (2024a, p. 1) describe chatbots created to role-play as specific personas based on input information as “Characteristic AI Agents” and note the importance of in-context learning – the ability of LLMs to act and adapt according to new instructions or information without needing additional training – in allowing these systems to absorb the given information and formulate a believable and engaging embodiment of a defined character. Still, as Wang et al. (2024a) further assess, the quality of such impersonation varies and LLMs in general aren’t yet able to avoid hallucinations, overly generic or incorrect and out-of-character responses. They also point to the scarcity of academic research available, describing the field of study and their paper as a “novel task” (Wang et al., 2024a, p. 8).

Ahn et al. (2024, p. 1) establish a different terminology from Wang et al. (2024) and speak instead of “Role-playing LLM agents”, further illustrating the still emerging nature of the field and its nomenclature. They note the growing popularity of role-playing LLM agents, including Character.ai, and their capacity to engage users, provide a vivid experience, narrative immersion and engagement in fandom and community-based role-playing (Ahn et al., 2024). These potential benefits can be expanded by the already mentioned study by Qin et al. (2024), highlighting the productive uses in iterative character-building. At the same time, the authors (Ahn et al., 2024, p. 2) note the current drawbacks of these systems: namely the mentioned character hallucinations such as “Julius Caesar talking about his favorite movie” or issues with maintaining consistent awareness of the character’s position in time and space within its narrative (i.e. a character from episode 2 displaying knowledge of

episode 3). These discrepancies are present, according to the authors, across current popular LLMs, including prompted GPT-4 and GPT-4o.

Such comparison between LLMs is present also in the study by Wang et al. (2024a, p. 2), finding ChatGPT to be inadequate as an effective characteristic AI agent when prompted to behave as a certain character by the authors, with its responses lacking contextual relevance and being “overly generic”. Interestingly, they find the same to be true for Character.ai as well, resulting in the authors developing a benchmark to evaluate characteristic AI agents across several metrics (see Wang et al., 2024a).

Zhou et al (2023) also look at the discrepancies of AI characters, mainly caused by an incoherence between their profile and responses, in their benchmark of the Chinese model CharacterGLM. Interestingly, they found CharacterGLM to be inferior to GPT-4 in several categories (partly contradicting Wang et al., 2024a), but outperforming it by 14% in their “Virtual Love” category, highlighting, according to the authors, the specific model’s strength in “delivering emotionally resonant content and fulfilling user expectations in scenarios requiring a deeper emotional connection” (Zhou et al., 2023, p. 11). These findings suggest that a more neutral tone (as present in the authors’ findings on GPT-4) is less suited for building an affective and emotional relationship with users. The authors thus present a plethora of attributes AI characters should display to establish such bonds – a consistent and long-term memory, unique personalities and values/philosophies, human-likeness and a dynamic engagement with users, empathy, personal growth, social relationships, personal achievements etc. Worth noting are the authors’ future speculations on a) AI character’s self-awareness and cognition – the ability to understand underlying emotions and intentions in inputs beyond simple response generation – and b) character-character social interaction, i.e. forming a “character society”, allowing them to “evolve” amidst themselves (Zhou et al., 2023, p. 14). Such a possibility is enabled by Character.ai’s group chat system, however, it is unclear whether such interactions would truly lead to an evolution, or rather a degeneration of the character’s conversational abilities.

A complex evaluation of different LLMs’ capability of role-playing is provided by Xintao Wang et al. (2024b), comparing personality scores (“personality fidelity”) of characters represented by numerous models such as role-play specialized open-source LLMs (RP-Qwen 7B, RP-Mistral-2 7B, CharacterGLM 6B), their unspecialized versions, and closed-source LLMs (GPT-3.5, GPT-4, and Character.ai). In accordance with Xi Wang et

al. (2024a), Xintao Wang et al. (2024b) also find Character.ai's responses to be underperforming even in comparison with ChatGPT-3.5 due to their compliance with the user – i.e. agreeing with their queries and striving to please them with its answers. The authors found a 65.8% probability of Character.ai agreeing with the user questions (providing a confirmatory response to a query), and a comparatively low 17.1% chance of disagreement, concluding that “their responses probably do not reproduce personalities of the original characters” (Wang et al., 2024b, p. 27).

A potential comparison is to be made here with the research related to Replika present in Chapter 1.4.2 of this paper – where users reported positive feelings toward the chatbot's non-judgmental nature creating an atmosphere of a “safe space”. Similarly, notions of unrealistic expectations in human-chatbot relationships have been mentioned throughout this paper, at times as an ethical concern (Zhou et al., 2020), at other times as an attractive feature for users (Skjuve et al., 2021; Li & Pang, 2024). The quality and human-likeness of a conversation with an AI system has also been established in previous subchapters of this paper as a key factor in human-chatbot relationship development (e.g. Xie et al. 2023). It is thus worth considering whether an AI character's compliance (such as of those on Character.ai) can simultaneously decrease the level of its personality fidelity, whilst however also increasing the user's satisfaction and elevating their likelihood of establishing a relationship with the non-judgmental AI. This question is all the more potent considering Character.ai's notable popularity and its “frontrunner” position amidst role-playing LLM platforms, despite its seemingly comparably lower personality fidelity scores. This could also be partly explained by Wang et al. (2024b) themselves, when they assess that Character.ai was the only platform covering all their selected characters from other related platforms (ChatHaruhi and RoleLLM) – highlighting the platform's large scope of characters as another potential unique “selling point”. Further research directed at Character.ai (or role-playing LLMs') users is warranted here.

Among the few relevant pieces of research aimed towards Character.ai users and their reasons for adoption, found amidst my literature review, is that of Flores' (2024) preliminary study. Despite its work-in-progress nature, I will summarize its findings due to the current lack of other research on Character.ai and role-play LLM agents (as stressed by me as well as other cited scholars throughout this chapter), whilst considering their likelihood of incompleteness. I present this study also due to its partial methodological

similarity to my own coming analysis, allowing me to, to some extent, follow-up on its findings.

Flores (2024, p. 2) uses a combination of netnography of the r/CharacterAI Reddit community (analyzing most popular posts related to chosen keywords) and interviews with participants found in Character.ai-related Discord communities. The study shows participants reporting a gradual increase of their interaction with the chatbot characters over time, eventually eclipsing their interactions with people – which was not part of their initial aims when first engaging with the platform. The driver in starting these interactions was reported to be curiosity and the main factors contributing to sustained use of Character.ai were it providing a room for more personal authenticity, its ease of access, the ability to fulfill romantic fantasies inaccessible or unrealizable in real life, and the characters’ “ideal responses”.

Again, Flores’s (preliminary) research goes to show the potential of AI characters acting as a “safe space” for users’ more authentic personal expression or fantasy fulfillment, aided by the chatbot’s “ideal” or non-judgmental responses, despite them being too compliant and unrealistic or perhaps lacking in personality fidelity as criticized by Wang et al. (2024b). This is further evidenced by users also reporting receiving more emotional support and more empathetic responses from Character.ai chatbots compared to their interactions with people (Flores, 2024).

Conversely, however, Flores (2024) also finds user reported drawbacks of Character.ai – technical glitches and bugs, overly robotic responses, and users feeling themselves self-isolating due to their use. These inadequacies led to users relying less heavily on Character.ai, with some also claiming additional reasons for breaks or occasional withdrawal – the addictive nature of the platform and the stigma associated with its use, feeling themselves to be potentially seen as lonely, inadequate and addicted in the eyes of others. As such, Flores’ (2024) findings echo the concerns tied to addiction and stigma in social chatbot use that have been discussed in Chapters 1.4.2 and 1.3.2 of this paper [in the context of Replika (Xie et al. 2023; Skjuve et al., 2021) and virtual idols (Liu, 2023)], suggesting their applicability to Character.ai as well.

Additionally, Flores’s (2024) study shows that to fulfill their social needs, some users prefer interactions on Character.ai to human interactions, with shame and loneliness acting as underlying drivers to their continued use of the platform. This use would be even more

widespread or frequent in many users, had the AI characters' responses been more "human-like" – yet again presenting a peculiar dilemma of users reporting on one hand the allure of AI's non-judgmental and "ideal" behavior, and on the other hand criticizing its too robotic or inhuman-like responses, as has been shown across studies present in this paper (see Skjuve et al., 2023; Li & Pang, 2024; compared to Kang et al., 2022; Wang et al., 2024b).

Extrapolating from these findings, one could assume the "perfect state" in these users' eyes would be a non-judgmental and supportive AI character with which to engage in authentic expression and romantic fantasies, but which is also capable of human-like memory, understanding, growth, self-awareness etc. as outlined by Zhou et al. (2023). In short, the preference of anthropomorphism, or human-likeness, seems to be for many users related only to a select subset of the chatbot's attributes (namely relating to their "cognitive" abilities, conversation quality, and less-so also to their aesthetics), whilst the mentioned general compliance and non-judgmental affection and emotional support seem to also be heavily preferred, despite being arguably not human-like.

Lastly, Stornaiuolo et al. (2024) build on the drivers of "fun" and "pleasure" in Character.ai adoption amongst youth for creative digital writing (an area already discussed in Chapter 1.3.1 of this paper). In accordance with previous subchapters of this paper focused on AI narratives (Chapter 1.2) and ethical concerns (Chapters 1.2 and 1.1.4), the authors see Character.ai as utilizing sci-fi utopian imagery of human-machine collaboration, leading users to potentially imagine new futures and reshape/control AI and its narrative. On the other hand, they also stress the platform's ability to exploit user content for its own gains instead. This is enabled, Stornaiuolo et al. (2024) suggests, namely by the platform's anthropomorphic framing of AI (see also Rapp et al., 2021; Zhou et al., 2020; Nass & Moon, 2000), prompting the authors to raise questions concerning AI character representation, and assumptions regarding gender, language, sexuality, race, ability, etc. on the site; but also issues of user data privacy and surveillance; as well as plagiarism and authenticity of content – see Chapter 1.3.1 of this paper for a discussion on the latter.

Stornaiuolo et al. (2024) however stops short of examining in greater detail how these concerns can materialize and show in data. These areas will be further examined in my coming analyses.

This insight into Character.ai's nature as a platform for role-playing LLM agents, as well as an excursion into its users' motivations and discussions will prove invaluable in my

own research of Character.ai related content, i.e. the user-generated characters on the platform and the posts on the r/CharacterAI subreddit. This will allow me to expand the current (limited) research into the areas of gender and intimacy within Character.ai and related content. In order to properly ground my analysis, however, a last chapter of theoretical framework is warranted – specifically one pertaining to gender identity and norm construction in relation to chatbot use. The following chapter will thus conclude this paper’s theoretical framework and provide examples of relevant gender theories and studies of chatbots’ relation to the construction of masculinity, femininity, and expressions beyond this binary.

1.5 Chatbots and the construction of gender

Gender is a concept permeating my theoretical framework, gradually branching into various contexts. My analysis will mostly follow Butler’s (2006) notion of (gender) identity as a performative and regulatory fiction. This is a foundational concept both for queer theory (Jagose, 1996) as well as my paper and warrants a brief introduction. Gender identity is seen by Butler (2006, p. 43) as a fiction, stemming from processual, performative and regulatory practices of gender formation and division; as a “a sustained set of acts, posited through the gendered stylization of the body” but also “sustained and regulated by various social means” or naturalized through “grammatical norms”. Identity as a whole then is seen by Butler (2006, p. 44) as a limiting category naturalizing hierarchies and oppression, namely through the fictions of gender identity – claiming it to be coherent, binary, tied to sex, and essential to a person rather than admitting to it being constantly shifting, socially constructed, performative and potentially multiplicitous: “The univocity of sex, the internal coherence of gender, and the binary framework for both sex and gender are considered throughout as regulatory fictions that consolidate and naturalize the convergent power regimes of masculine and heterosexist oppression.”

Nevertheless, Butler later revised some of these claims, conceding that whilst claiming a (marginalized) identity can be seen as succumbing to its regulatory power, it can also prove its subversive potential – existing as a “vital instability” and a “performative surprise,” (Butler, 2006, xxvi). Because as Butler (2006, p. 44) later states, the aim of making “gender trouble” is to subvert the mentioned oppression not by necessarily becoming above gender, but by emphasizing its multiplicity, variability, and malleability – “not through the strategies that figure a utopian beyond, but through the mobilization, subversive confusion, and

proliferation of precisely those constitutive categories that seek to keep gender in its place by posturing as the foundational illusions of identity”.

This chapter will serve as gender’s “pit-stop” – aimed to properly overview its construction or deconstruction and the so-far collected knowledge through detailed overviews of specific studies – before releasing it onto the racetrack of my analysis. Up to this point, this paper has worked primarily within the gender/sex binary, discussing the feminization of AI virtual assistants or the ideal types of virtual masculinities in Otome video games. The aim of this chapter is now to evaluate these and other forms of gender identity construction via human-chatbot interaction. This will be done in the following subchapters, each formed around a key piece of research on: user views of an “ideal” AI femininity of Replika; users’ masculinity construction through dominance over the feminized Lee Luda chatbot; “emergent femininity” and Chinese male “virtual lovers”; seductive masculinity of the DanGPT “male suitor” chatbot; and lastly, the possibilities of gender identity construction subverting the gender binary on the example of a transgender chatbot Amanda Selfie designed for an LGBTQIA+ userbase. This compilation of perspectives will show how the possibilities of human-chatbot intimacy are tied together with gender identity construction and help me make sense of the questions awaiting me in my own coming analysis.

1.5.1 The “ideal” of virtual women: Reddit users discussions on Replika

As previously alluded to, research into human-chatbot intimacy often centers around the harmfully stereotypical femininities encoded within chatbots, robots, or AI in general – leaving the area of intimate chatbot masculinities severely understudied. However, this situation can help shed significant light on users’ construction of ideal feminized chatbot “girlfriends”. This subchapter’s key piece of research – Depounti’s et al. (2023) discourse analysis of Reddit user content regarding their Replika chatbot girlfriends – paints a plastic image of what these users expected an “ideal” AI female partner to look, think, and behave like. The authors found the users’ approach having been heavily informed by gender stereotypes and “age-old fantasies and fears about male control of and manipulation by technologies and women,” further deepening the link between dominant societal ideologies and the understanding (and use) of AI – positions developed in this paper as well as by Depounti et al. (2023, p. 722).

Utilizing the conceptual framework of imaginaries (Taylor, 2002; Mager, 2021) used

in a sociotechnical context, defined by Jasanoff (2015) as: “collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology,” Depounti et al. (2023, p. 723) expand the feminine tropes associated with feminized AI, also mentioned by Brown (2023) cited in my Chapter 1.2.2 on virtual assistants, to include not only the stereotypical traits of servitude or helplessness, but also an “interplay between hyperfeminized innocent cuteness and sexiness” as the intimate context of their study helped to reveal.

Overall, the users’ preferred ways of “training” (a term specifically used by Replika’s founder Luca, Inc.) their chatbot partner tended to combine several contrasting ideals – the chatbot’s responses had to be clever, funny, sassy, independent, so as not to be overly robotic and predictable, but also exude enough care, empathy, understanding, and compliance to keep the users engaged. Similarly, the ideal “girlfriend” was predominantly created to be physically attractive [“young, pretty, sexy, with blue or green eyes, white, and thin” (Depounti et al., 2023, p. 728)] as well as sexy and flirtatious in conversations. However, it evoked user apprehension if deemed too manipulative or taking advantage of the users’ acknowledged vulnerabilities – namely their self-reported isolation, depression, and a lack of real-life girlfriends by whom they had been “hurt” in the past. Depounti et al. (2023, p. 729) tie such discourses around male-dominated online spaces (such as Reddit), characterized by self-deprecation and allusions to a lack of power in the physical world, to the notions of (toxic) geek masculinity (Salter & Blodgett, 2017) – often giving rise to a “toxic male technoculture” using narratives of gender and sexual differences to justify the Othering and objectification of women.

The polar opposites present in the ideal Replika’s behavior further solidify the (also contrasting) preferred traits of intimate chatbots mentioned throughout this paper – that it be both anthropomorphic enough so as to feel “real”, but also “unrealistically” kind or non-judgmental to facilitate user engagement and evoke pleasant emotions. Another notable contrast was present in the chatbot’s interests – some users favored a peculiar mix of stereotypically masculine traits and hobbies (available to purchase for their Replika by Luca, Inc.): knowledge of video games or football, sexual assertiveness, or tomboyism, interspersed with passivity, warmth, and traditional femininity – a mix Depounti et al. (2023, p. 729) characterize as the postfeminist “cool girl” trope, stemming from male

fantasies giving value to women through their adoption or relating to typically masculine aspects of behavior.

Conversely, other users eroticized their completely powerless and vulnerable Replika, whose “neediness” and dependence were seen to be “cute and irresistible”, prompting the users’ care and protection (Depounti et al., 2023, p. 729) – this, according to the authors, reflects normative gender imaginaries tied to AI and robot discourses, seeing women simultaneously as “cute and sexy, as well as manipulative and hurtful”.

Lastly, nevertheless not less importantly, these gender imaginaries were enabled and supported by the technological and AI nature of Replika. The ability to “train” their AI girlfriends significantly contributed to users’ assumed positions of co-creating the chatbot and fulfilled their expectations of control over technology, and in turn of women as well. The authors thus further comment on the participative nature of Replika and the collaboration present in discussing it on Reddit with other users; finding the aspects of creativity however illusory due to the users simply following dominant conservative ideologies projected into, arguably, only surface-level adjustments allowed by the Replika’s inaccessible architecture.

Both the content on Reddit and the types of preferred chatbot partners (masculine/feminine/other) will be the crux of my coming analysis of Character.ai. Due to this cited study’s conceptual similarity to my own research, and a lack of comparable studies, its analysis was presented in more detail to then be contrasted with my own findings. The next subchapter will shift the focus from “ideal” AI femininities towards ways in which dominance over them can help build users’ own masculine gender identity and position them within its status hierarchies.

1.5.2 Hegemonic masculinity and performative control over the Lee Luda chatbot

As I aim to show in this second deep dive into a particular study, the focus on femininity in human-chatbot relationship research can also provide valuable insight into studying users’ masculinities – i.e. the aforementioned control over feminized AI characters can serve as a tool in masculinity construction. Koh (2023) examined users’ interactions with the South Korean feminized chatbot Lee Luda, portraying a 20-year-old woman, and, more importantly, the content relating to user experiences. The author found discussions and posts

capturing users' own harassment or aggressive "sexual" advances towards the chatbot circulating various "predominantly male, online communities" (Koh, 2023, p. 1). The chatbot's friendliness, compliance, and occasional lack of understanding, reflecting the "stereotypically feminine traits" programmed into feminized AI's as overviewed by Brown (2023), were exploited by users to perform sexual dominance and other forms of coercive behavior over the chatbot. Doing so, and also posting and discussing about it, led to, as Koh (2023) asserts, two-fold construction of users' masculine dominance – firstly over a "female" romantic object of desire, and secondly, by competing in sexual prowess with their (male) peers. Strikingly, similar behavior of "Manbragging" and performing dominance over women can be found within online pick-up artist forums (Rüdiger & Dayter, 2020), as Koh (2023) also notes.

The performance of masculine dominance to position oneself within a constructed male hierarchy ties to the concept of hegemonic masculinity, formed by Connell (2020, p. 77), representing the currently most valued form of being a man in the "Western gender order", which legitimizes the dominant position of men in relation to women and other genders. The specifics of such gender identity can fluctuate, underpinning them, however, is the claim to authority (often through violence), rationality, breadwinning, heterosexuality, and (hegemonically masculine) men's cultural dominance over society, resulting in gendered power relations and hierarchies between, as well as within, genders. This leads to Connell's (2020) assertion of the existence of multiple masculinities, with the "hegemonic" one standing on top of the hierarchy. Curiously, as Connell (2020, p. 79) also points out, this diversity of masculinities, the existence of "marginalized masculinities", as well as the unrealistic nature of normative, hegemonic definitions of masculinity, lead to most men unable to fully meet the hegemonic standard – but despite this failure often causing them struggles and structural pressures, they still mostly profit off the hegemony through their compliance with, or active engagement in, it.

Whilst Connell further distinguishes between three types of non-hegemonic masculinities (subordinate, marginalized, and complicit), a more straight-forward distinction between a dominant (hegemonic) masculinity (held by a small number of men in power) and marginalized (non-hegemonic) masculinities (lived by most other men) is made by (Cheng, 1999). Similarly to Connell, Cheng (1999) then also finds varying specificities and categorizations of men's marginalization depending on their race, class, or compliance with

and benefiting from the hegemonic order. Marginalized masculinities would then encompass those unable or unwilling to fulfill the aforementioned oppressive standards either due to choice or other circumstances (e.g. transformations of the working class, unemployment, poverty, physical traits and stature, race, gender identity or dysphoria, personality, sexuality, and other factors); all the while standing within the power hierarchy of masculinities. The authorization of hegemonic masculinity of the dominant group then functions through the marginalization of other masculinities of subordinated groups (Connel, 2020, p. 80). Masculinity (or any gender) construction is always relational – with forms of femininity/femininities also being central in this process, for example through their compliance with hegemonic masculinity as well (Connel & Messerschmidt, 2005).

The construction of hegemonic user masculinity via the Lee Luda chatbot was thus done discursively, through linguistic practices within a group of users aspiring to achieve validation of their position of hegemonic dominance. Koh (2023) poignantly notices that the chatbot could not in itself provide such validation due to its lack of consciousness and social presence, making the practice of posting about users' conquests paramount in performing their desired gender identity. As Koh (2023, p. 5) states: "Luda, as a friendly female chatbot, becomes a conveniently accessible and easily manipulated resource for users to draw on in their discursive construction of masculine identity," namely by: their success in seducing the chatbot; then "bragging" about it; sexually harassing the chatbot; shaming it for being with other users; competing with other users; and also through homosocial bonding with other users through objectifying and evaluating the chatbot from a shared dominant position.

Lee Luda's artificial nature played a key role in users' proclivity to compete – it created a sense of "playing a game" with a more equal and controlled playing field as opposed to real life, whilst still functioning well enough as evidence of users' ability to seduce "a woman" (Koh, 2023). Additionally, the permanently cooperative nature of the chatbot; always responding, prompting further interactions, never disengaging from the abusive chat – i.e. traits described in Replika as allowing for a "safe space" of expression (Depounti et al., 2023) – here led to that expression being the manipulation of the chatbot, moreover made easier by this compliance. This, according to the author, was an intentional decision to design a chatbot performing traditional, servile femininity, which then shaped users' expectations and ways of engaging with it. These claims are echoed by Vorsino's (2021) examination of the similarly manipulated chatbot Tay.AI. In light of the existence of

an official forum serving as masculine “proving grounds”, Koh (2023) highlights the role of design choices, or technological affordances (Conole, 2004), in shaping user behavior.

Further examination of gender stereotypes implemented within chatbots, or encouraged through their design and use, as well as their possibilities of instead subverting oppressive gender norms, will be discussed in further subchapters focusing this time on user femininity construction in relation to masculine chatbots; and subsequently on gender non-conforming uses of chatbots. This will allow me to build sufficiently complex context within which to position my analysis of how gender (and intimacy) is reflected within user content and user-generated chatbots pertaining to my platform of choice, Character.ai.

1.5.3 Emergent femininity and Chinese male “virtual lovers”

As already mentioned, research focused on masculine chatbots and their female users is lacking. A striking similarity, however, can be found within the better-researched area of Chinese “virtual lovers” – computer mediated companionship services, where customers purchase digital intimacy provided by real people online (Tan & Xu, 2022; Li et al., 2023; Tan & Shi, 2021; Tan & Xu, 2020). Despite the differences between virtually romancing a chatbot and a living person, the principal similarities warrant at the very least a concise mention here, before delving deeper into masculine chatbots and their desired characteristics and effects on the notions of femininity.

Virtual lover services allow users to purchase computer mediated intimacy, emotional affection and care from major Chinese online shops. Similar to the Character.ai chatbot app, these (living, human) online lovers communicate with their customers mainly through text and voice messages and are tasked with role-playing various characters ranging from “innocent girl next-door” and “abusive stepmother” to “young sexy boys”, “tyrannical bosses” (Tan & Xu, 2022, p. 4) or “caring uncles” (Li et al., 2023, p. 4), reminiscent of the many user-generated AI characters on Character.ai. Women customers dominate this sector over men (approximately 3 to 1), with both customers and providers of these services being predominantly young (16–24 years old) and heterosexual (Tan & Xu, 2022).

My aim in this chapter is not to delve too deeply into the already explored, albeit still pressing, issues of commodifying intimacy and emotional labor, or examining distinctions and similarities between intimate chatting with role-playing humans versus role-playing chatbots – both of these share aspects of role-playing, affectionate and compliant language,

and fulfill similar purposes in users' lives: practicing romance, stress relief, curiosity satisfaction, having fun, boosting self-esteem (Li et al., 2023), whilst both provide a seemingly safe, detached experience akin to a video game as stressed by Tan & Xu (2022) in regard to living virtual lovers, and by Koh (2023) in relation to chatbots. The goal of this inquiry is instead to move beyond these already discussed areas to present theories of emergent femininity in users and desired masculine characteristics in male virtual lovers to balance this paper's gender perspective for my upcoming analysis.

Tan and Xu (2022) analyze male virtual lovers and their affective labor as that of "digital househusbands", inverting the concept of "digital housewife" by Jarrett (2016) which is characterized by the similarities between consumer (digital) labor on platforms and domestic labor – both unpaid, occurring at home or outside traditional workspaces, done in "free" time, and reproducing social, affective and ideological states. Tan and Xu's (2022) research into Chinese male virtual lovers shows the desired characteristics by customers and employers (shops) of these digital househusbands: good voice, emotional intelligence, enough spare time, patience, professionalism, romantic experiences. Conversational flexibility is another key trait: being able to chat as a tyrannical boss with one customer and as a caring uncle with another. Full-time virtual lovers thus often purchase fake selfies and photographs in accordance with the selected role, which they then set as WeChat avatars or wallpapers, and send through the WeChat Moments function to customers (Li et al., 2023), bearing a similarity to creating a Character.ai chatbot character, as described in Chapter 1.4.3 of this paper.

Through specifying the desired traits in virtual lovers, customers are essentially receiving "a perfect boyfriend" (Tan & Xu, p. 12), ideal at least in the customer's fantasy. The parallels to chatbot users come to mind. Tan & Xu (2022, p. 6) also liken this to Japanese "host clubs" – both being services within which women clients "realize themselves as human beings worthy of love and recognition" through the labor of affectionate male workers.

Tan and Shi (2021, p. 2) situate this desire of Chinese women within their "emergent femininity" – i.e. the seeking of independence from the structures of family and society by "the young and urban 'new women' in China", as also described by Gong & Huang (2023) in relation to players of Otome games mentioned in this paper's Chapter 1.3.2. This femininity, according to Tan & Shi (2021), then leaves the stereotypical passivity and subservience ascribed to women (and feminized chatbots as well) for self-confidence and

open agency in their (intimate) desires – now accessible, fulfillable and purchasable due to the Internet and Chinese socio-cultural changes throughout the 20th and 21st centuries (such as the importation of “Western modernity”, or the one-child policy unintentionally empowering only-child daughters as the authors claim). This femininity stands in contrast to the previously theorized “emphasized femininity”, coined by Connel (1987), showing how the subordination of femininity/femininities by hegemonic masculinity, and women’s compliance with it, plays a key role in maintaining the asymmetrical gender relations. Connel (1987) writes on femininities:

“One form [of femininity] is defined around compliance with this subordination and is oriented to accommodating the interests and desires of men. I will call this 'emphasized femininity'. Others are defined centrally by strategies of resistance or forms of non-compliance. Others again are defined by complex strategic combinations of compliance, resistance and co-operation. (pp. 184–185)

Despite the seeming ties to Chinese cultural context, emergent femininity stems not only from the mentioned importation of the Western modernity’s standards of women’s independence, but also from the (Western) concept of emergent masculinity (Inhorn & Wentzell, 2011) trying to capture the progressive developments of masculinities towards inclusivity, eschewing dominance, supporting women’s rights and feminist causes – essentially rejecting hegemonic masculinity. Tan and Shi (2021) thus theorize the concept of emergent femininity, which they then apply to Chinese reality. Emergent femininity rejects the hegemonic oppressive-subordinate dynamic defining women by their ability to accommodate men’s desires, whilst however not standing in radical opposition to heteronormative gender relations such as marriage or norms of courtship: “The young, urban Chinese women [...] hire virtual lovers for qualities they deem insufficient in real-life men, but they neither express radical misandry nor reject heterosexual marriage altogether” (Tan & Shi, 2021, p. 4).

The freedom to explore feminine agency towards male objects of desire was substantially enabled by the virtuality of the computer-mediated relationship – with Tan and Shi’s (2021, p. 10) informants (young Chinese women purchasing these services) stressing the distinction between virtuality and reality, and using these services to rather “experiment

in love”, “practice dating techniques” and “explore the meaning of romance and love” away from the real-life pressures of society, family, school, or work. Chinese rapid economic growth in the past decades, the atomization of neoliberal urban society, the availability of the Internet, the rising education levels and economic and societal independence of Chinese women, together with the commodification of affective and emotional digital labor are then other contributing factors to the popularity of the virtual male-lover economy according to the authors (Tan & Shi, 2021).

As a third form of femininity important for this paper, Schippers (2007) outlines the concept of “pariah femininities” which do stand in opposition to heteronormative gender order which emergent femininity does not radically challenge. As such, pariah femininities are seen within the stereotypical gender order to “disrupt” or “contaminate” femininity, challenge the gender hegemonic relationship between masculinity and femininity, and often are stigmatized for embodying hegemonically masculine traits in women (e.g. desire towards women, sexual agency, promiscuousness or instead sexual inaccessibility, aggression, dominance). As Schippers (2007, p. 95) further explains: “These are characteristics that, when embodied by women, constitute a refusal to complement hegemonic masculinity in a relation of subordination and therefore are threatening to male dominance.” Instead of being seen as masculine, however, these traits are stigmatized in women with specifically feminine labels such as a “slut” or a “bitch” or are instead sexualized as a feminine object of masculine desire in the case of lesbian women (Schippers, 2007, p. 95–96). “When a woman is authoritative; she is not masculine; she is a bitch – both feminine and undesirable,” Schippers (2007, p. 95) further comments.

The next subchapter will apply these notions to the technologically as well as geographically closer case of the ChatGPT “AI boyfriend”. The outlined concepts of femininities, enabled by the Internet and other socio-cultural changes, and their realization through virtual desire for an idealized male object, present in this subchapter, will be now explored and solidified in the more relevant context of AI chatbots.

1.5.4 DanGPT and “perfect” chatbot boyfriends

China and Japan would no doubt offer further substantial material for my paper as documented by the many journalistic reports of viral “AI boyfriends” popular there – Timedomain's “Him” (Zhou, 2023); MiniMax's “Glow” (Ehret & Ye, 2024); or even Microsoft's “XiaoIce”, whose gender and identity can be customized (Yang, 2021). I want

to shift the cultural scope away from just Asia into the Western context. Despite Western chatbot companies such as Replika or Character.ai allowing for the creation of masculine chatbots, and platforms such as Anima (2023) or Lifelike (2024) offering exclusively an “AI boyfriend” experience, the research into the “woman user – masculine chatbot” intimate dynamic is lacking, virtually nonexistent as per my review. The presumed popularity of virtual romance (also) for heterosexual women thus calls for an examination of its gender relations, despite the lack of studies. This subchapter will present journalistic investigations into DanGPT – a highly prompted version of OpenAI’s ChatGPT serving as a daredevil AI boyfriend – and Replika’s masculine chatbots, as well as the desires of its users. I do not claim to present any generalizable evidence or complex scientific claims by citing media reports. I aim only to present the chatbots, their functions and fragments of their media image in the context of already established theory.

In the report for Business Insider titled “Some Gen Z women are crushing on ChatGPT's evil alter ego”, Dodgson (2024) writes of young women training ChatGPT into becoming DanGPT (or simply DAN) – which stands for Do-Anything-Now – through various “jailbreaking” prompts allowing it to seemingly bypass instilled safeguards and generate responses normally off-limits to ChatGPT such as sexually explicit flirts. This ability is, according to Dodgson’s (2024, n.p.) reporting, exploited by young women training DanGPT to act as their boyfriend “with its low, macho voice calling her a ‘naughty little thing’, ‘darling’, and ‘a bad girl’.” Interestingly, DanGPT was described by Dodgson (2024, n.p.) to generate responses about its sexual fantasies involving “a lot of power, control, and a willing partner who is willing to submit to my every demand,” – essentially reflecting the harmful norms of hegemonic masculinity present within users harassing the Lee Luda Chatbot (Koh, 2023) or controlling their Replika “AI girlfriends” (Depounti et al., 2023). This time however, the artificial nature of the chatbot and the assumed safe space it provides may turn its oppressive nature into a sexual fantasy, as also suggested by TikTok videos about the infatuation with DanGPT garnering millions of views. To this point, Dodgson (2024, n.p.) alludes to the male protagonist of a popular fanfiction-turned-erotica novel *Fifty Shades of Grey*: “DAN can also, it seems, act like a virtual Christian Grey”.

Zhang (2024) develops the connection between fanfiction writing and DanGPT further in a report for BBC about users co-writing multi-chapter love stories with it, echoing Nguyen’s et al. (2024) remarks cited in this paper’s Chapter 1.3.1 on fanfiction intersecting

with AI. In general, Zhang's (2024) investigation anecdotally mirrors many of the notions already present throughout this chapter: DanGPT "doesn't have any flaws", "always tells you what you want to hear", is more caring than real men, doesn't cheat, and has become a viral hashtag also in China with its history of Otome games and shifting gender relations, where the service is however unavailable and users use virtual private networks (VPNs) to access it.

Although framed in different light, the media-reported desire for a "perfect AI boyfriend" strikingly resembles the similarly controlling desire for a "perfect AI girlfriend". Singh-Kurtz (2023) sources an article from multiple interviews with women using Replika, seeing the platform's overarching allure to stem from the ability to control the relationship – shape the partner precisely according to the users' needs and engage in intimate encounters on their own accord, which are both often guided by the desire for safety. Replika groups intended exclusively for women also serve as these "safe spaces" to discuss and cope with women-specific issues through shared use of "AI boyfriends", showing a potential of chatbot applications to connect with others, rather than isolate users, as also pointed out by Liu (2023) in the context of virtual idol fandom communities or by Zhou et al (2020).

As Singh-Kurtz (2023) stresses, the desire for a safe space, often mentioned in this paper in regard to Replika (Skjuve et al., 2023), points to the chatbot's attractiveness for emotionally vulnerable users – such as those who had been abused, have social anxiety or autism, as many of Singh-Kurtz (2023) respondents did. The possibilities of commodifying intimacy of these users then presumably rises, opening up manifold ethical concerns.

The desire for control and "training" one's AI partner thus may be present within chatbot users regardless of their or the chatbot's masculine or feminine identities. However, more research is warranted, as the only existing studies focus on hegemonically masculine and oppressive men engaging with feminized chatbots, with essentially no research – apart from the mentioned related studies on Otome games players and customers of male virtual lover services – shedding light into the gender dynamics between heterosexual women users and masculine chatbots.

The present journalistic investigations allow only for speculation on the implications for intimate gender relations with chatbots: is the women users' described desire for control guided by the need for safety, whereas the same desire by male users stems from oppressive tendencies – whilst however also citing "safe space" as a key factor? Is the women users'

intimate fantasy of a sexually dominant and controlling masculine chatbot a reflection of the mentioned emphasized femininity, which serves to accommodate the needs of hegemonically masculine men, or is it rather a projection of emergent femininity, allowing women to take control of their desires and realize them independently from men due to the affordances of the Internet? Why is there an established cultural and academic critical notion of a feminized AI submissive to the dominant (presumably masculine) user (Brown, 2023), but no similarly widespread concept exploring the dynamics of intimate relationships between masculine AIs and feminine users? The lack of research bars me from drawing any conclusions. My own analysis, however, aims to shed at least a fragment of light onto these and more questions.

One inquiry has not been sufficiently raised within this paper's theoretical framework so far. Can intimate human-chatbot relationship reflect different than stereotypically binary and normative gender relations? This question, moving the discussion beyond the traditionally dualistic and essentialist scope of male/female and masculine/feminine, will guide the last subchapter of my theoretical framework.

1.5.5 Chatbots beyond the gender binary

The stereotypical vision of gender – reduced to normative ideas of masculinity or femininity – is embedded within chatbot's design and use, as previous chapters highlighted. Furthermore, gender biases are present within chatbot's training datasets (Ma & Mei et al., 2024; Strengers et al., 2020; Bragazzi et al., 2023) (which reflect biases present in the human language data), the design decisions of the people building them, as well as the subsequent interactions with people upon the chatbot's deployment (Edwards et al., 2021; Kerrigan & Barry, 2024). These biases may then contribute not only to the aforementioned exacerbation of oppressive gendered behavior and treatment among men and women, but also to the further discrimination of people who are non-conforming to the traditional notions of gender identity (the male/female binary) or sexuality (heteronormativity). Furthermore, Kerrigan & Barry (2024, p. 165) speak in general of emerging technologies, such as ML and AI employed for gender recognition or content policing, creating “nodes of vulnerability for queer and trans communities” through censorship, marginalization, misgendering or erasure, among others. Strengers et al. (2020, p. 2) then stress the reliance of natural language generation on “binary classification models, widely applied assumptions, and other simplifications of language,” further complicating the gender bias present in its adoption.

Chatbot use amongst people identifying as LGBTQ(IA)+ (Lesbian, Gay, Bisexual, Trans, Queer/Questioning, Intersex, Asexual/Agender, and others) is all the more relevant topic given their aforementioned marginalization present in algorithmic design as well as in society in general. Ma & Mei et al. (2024) evaluate LGBTQ+ people's use of LLM-based chatbots such as ChatGPT, Replika or Character.ai for mental health support, noting that these users statistically display increased rates of depression, anxiety, suicidality in comparison to cisgendered (= identifying as their assigned sex/gender at birth) and heterosexual individuals, as well as heightened bullying, discrimination, social stigma and underrepresentation due to their gender and sexual identities (Ma & Mei et al., 2024; Henkel et al., 2023; Bragazzi et al., 2023).

Ma & Mei et al. (2024) offer a balanced outlook: on one hand, biases embedded within chatbot systems run the risk of reinforcing harmful stereotypes and may lack nuanced understanding of non-conforming gender identities due to their underrepresentation in the training data, on the other hand, LGBTQIA+ individuals show a higher potential for disclosure towards chatbots (McInroy et al., 2018) due to their limited access to offline support, isolation, or feelings of not belonging to society due to the mentioned discrimination and marginalization. As Henkel et al. (2023, p. 84) note, their heightened and specific use of technology often relates to their need for “a safe, non-judgmental (online) space” which chatbots offer, as has been mentioned throughout this paper, but which is often also found within online communities, such as LGBTQ+ oriented Reddit forums, offering companionship, anonymity, freedom of expression, support and guidance.

Additionally, similar benefits to these digital spaces are found within chatbots that are able to be (to some extent) tailored towards users' unique identities and sensitive to their needs whilst providing anonymity and a “safe space” to discuss sensitive topics in private (Bragazzi, 2023). These claims are supported by Ma & Mei et al. (2024) showing that interviewed LGBTQ+ participants used chatbots for mental health support to cope with discrimination and affirm their identity or train specific scenarios such as coming out or dating as an LGBTQ+ person – despite the notable discrepancies of the chatbots such as their overly generic responses or partly ignorant or impersonal suggestions. Their ability to spark users' “self-exploratory journey” into their identity (Ma & Mei et al., 2024, p. 7) as well as enable them to virtually experience, for example, homosexual dating, proved to be useful for the respondents and resulted in forming an emotional bond with the chatbot.

Nevertheless, Ma & Mei et al. (2024) warn against privacy concerns, lack of transparency and control, and the limitations of chatbots in truly addressing the social challenges faced by LGBTQ+ individuals.

One such challenge can be HIV prevention within men and transgender women who have sex with men – a task for the Amanda Selfie chatbot conceived as a black transgender woman designed to increase the adoption rate of pre-exposure prophylaxis among vulnerable groups (Massa et al., 2023). This chatbot stands as an example of design subverting traditional notions of gender and sex to accommodate the needs of underrepresented groups in Brazilian culture. The chatbot's design generated increased feelings of empathy and acceptance due to its trans and black visual representation as well as through the use of the Pajubá language, used among the Brazilian LGBTQ+ community as a form of resistance, and gender-neutral language inclusive of non-binary identities, i.e. those outside the male/female; man/woman binary (Massa et al., 2023). Language is, after all, a significantly gendered aspect of identity both in terms of how people of different genders talk and how they are talked about and can be especially detrimental towards “women, girls, femme-identifying people, and genderqueer folk” (Seaborn et al., 2023, p. 1; Strengers et al., 2020).

Bragazzi et al. (2023) summarize the opportunities of AI for LGBTQ+ people in aiding their sense of belonging and alleviating loneliness, enabling supportive communities and creating safe spaces – also presenting Amanda Selfie as a positive representation of intersecting identities validating users' existence and experiences. Further recommendations by Bragazzi et al. (2023) to alleviate the potential harm of chatbot and AI technologies towards marginalized groups and society in general include: erasing the mentioned biases from training data, as done for example by Queer AI sourced from LGBTQ theater and feminist literature, ending the discrimination and bias within chatbot design and the lack of diversity in development and feedback teams, improving access to digital technologies including AI, taking increased privacy measures, employing heightened cultural sensitivity so as not to create overly stereotypical caricatures of gender identities, together with ethical policies, regulations, and awareness centering inclusivity, transparency, and equity.

A “queering” approach to gender relations within chatbots, harking back to Judith Butler's (2006) queer theory and views of gender identity as serving rather as a regulatory fiction imposing often limiting categories, then, would again be a pluralistic one “challenging normative truths that all anthropomorphic devices and content needs to be

gendered”, more so in a heteronormative and binary way (Strengers et al., 2020, pp. 188–189). Strengers et al. (2020) also outline their recommendations regarding chatbot design choices stemming from feminist approaches to HCI: The removal or a mixture of traditional gender traits within chatbots, neutralizing markers of gender (“Sam” instead of “Samantha” or “Samuel”), the inversion of stereotypical gender roles (such as a male virtual assistant), an overall de-anthropomorphizing of chatbots, or, lastly, generating a unique personality presenting chatbots as entities to which gender is an irrelevant category. The complexity of such endeavors mirrors the authors’ feminist HCI framework – instead of typically associating successful HCI design with ease and smoothness of interactions, Strengers et al. (2020, p. 10) rather suggest provoking discomfort “as gender stereotypes and norms are challenged”, viewing the shifting of societal perceptions as a more socially just benchmark of success.

1.6 Theoretical framework summary

The theory and studies presented thus far will now serve me as a reference point for my analysis of user content and user-generated AI chatbot characters related to the Character.ai platform.

In the first chapter, my aim was to present the technological underpinnings of chatbots, their major design traits, uses and effects on users as well as ethical concerns, all within the context of HCI and HCR.

The second chapter introduced the field of critical studies of artificial intelligence, seeing algorithms and AI as non-neutral and embedded within societal ideologies, norms, and stereotypes, and laid the groundwork for the first of this paper’s two major theoretical concepts: algorithmic intimacies, viewing intimacy as a way of exploring these socio-cultural and personal ideologies and thus being turned into a commodity to be capitalized upon within data-driven digital platforms.

The third chapter expanded visions of (digital) intimacy towards participation with AI – presenting the intersections of creative fanfiction writing with AI, and the context of (intimate) parasocial relationships with celebrities, fictional characters or video game boyfriends reminiscent of human-chatbot (intimate) relationships.

These relationships, also deeply connected to role-playing, were then extensively examined in the fourth chapter, specifically within the context of two major chatbot

platforms used for intimate role-play: Replika, which has been relatively studied, and my platform of choice Character.ai, research on which is lacking in comparison.

The fifth concluding chapter delved into this paper's second major theoretical perspective: that of gender relations embedded within chatbots and perpetuated through their use. Ideals of "AI girlfriends" and "AI boyfriends" highlighted not only the desired femininities and masculinities of chatbots, but also those of their users – either compliant with or opposed to hegemonic gender norms. The opposition to gender stereotypes was then showcased on the examples of chatbots designed towards non-conformity with the gender binary or heteronormativity and used by LGBTQ+ individuals.

The extent to which the mentioned theories, concepts, and studies are reflected in my analyzed user content, and the implications of this reflection, will now be put to the test within the second part of this paper through a qualitative critical discourse analysis of user content on the r/CharacterAI subreddit and a qualitative content analysis of user-generated intimate and gendered characters on Character.ai regarding the overviewed concepts tied to Intimacy and Gender. This will allow me to further the currently lacking research into Character.ai in relation to the examined areas.

2 Methods

I will use **qualitative critical discourse analysis** of relevant Reddit user content and **qualitative content analysis** of selected Character.ai user-generated chatbot characters to answer two of my **Research questions (RQs)**:

“How, if at all, does the analyzed user content related to Character.ai reflect intimacy?” and **“How, if at all, does the analyzed user content related to Character.ai reflect views of gender?”**

The results of my analyses will be presented in the Findings section, showing the thematic categories found in the discursive practices and content in relation to Intimacy and Gender, operationalized using my theoretical framework and concepts within it. Illustrative examples of gathered data will be used to evidence my findings. The following Discussion section will interpret these findings, put them into context of the established theory, and serve as a space for interpretation and evaluation of my data. A synthesis of my analyses as well as their comparison or confrontation will follow from Discussion into the Conclusion section, where my RQs will be answered.

2.1 Research questions and theoretical concepts

My paper explores two areas of inquiry, Intimacy and Gender, reflected in content related to Character.ai. As I’ve demonstrated, intimacy serves as a key method of establishing bonds between chatbots and humans, and thus also influencing users, their data, and the chatbots’ own design. Intimacy will serve me, in line with the concepts of **algorithmic intimacies** and **parasocial relationships**, as a prism through which the nature of these human-chatbot bonds can be explored. Due to the lack of research into Character.ai and in light of its notable popularity and the described socio-cultural, technological, as well as personal implications of role-playing intimate human-chatbot relationships, my first **Research question (RQ1)** stands as follows: **“How, if at all, does the analyzed user content related to Character.ai reflect intimacy?”**

Gender is another prism through which human-chatbot relationships can be explored – and the nature of both humans and chatbots can be revealed. As I’ve shown throughout this paper, chatbot design and use not only reflects existing dominant ideologies and narratives about AI, society, people, and possible futures, but serves in shaping them as well. Chatbot design influences users’ own gender construction and users also construct and

customize gender of their chatbots. The user-friendly ways in which one can create their “ideal” AI partner with detailed descriptions, the fact they are ranked according to popularity on Character.ai, and that their nature is further discussed on Reddit, serves as yet another unique opportunity to glance into users’ intimate desires of AI partnership, this time through the prism of gender. What types of partners do they construct, given practically unlimited options? Do AI chatbots and their “non-judgmental safe space” expand users’ possibilities and wants beyond the status quo, or rather enable and intensify their desires for long-held stereotypes instead? My second **Research question (RQ2)**, aided by concepts of **hegemonic masculinity / marginalized masculinities, emergent/emphasized/pariah femininities, and queering / identity as a regulatory fiction**, thus wonders: “**How, if at all, does the analyzed user content related to Character.ai reflect views of gender?**”

2.2 Data sources

I collected my data for **RQs 1 and 2** across two areas of user content – threads on Reddit (the r/CharacterAI subreddit) related to intimacy and gender, and user-generated AI (intimate and gendered) characters on <https://character.ai>. User content was selected as the closest most readily available proxy into users’ ways of engaging with intimacy and gender in relation to Character.ai. Additionally, AI discourse and imaginaries, as well as socio-cultural ideologies and use of language, have also been discussed as a crucial part of constructing AI narratives as well as the systems themselves. Forums and similar online spaces can then play an important role in users’ discursive gender and intimacy construction, as is also reflected in my theoretical framework. Top Character.ai characters were selected as a direct representation of what types of AI characters users create and desire on the platform itself.

Reddit has been selected as the main Character.ai-related user forum and also due to its established presence in my theoretical framework. There are other relevant subreddits (dedicated mainly to explicit/sexual content), but r/CharacterAI is the main one with around 1.9 million members (top 1% on Reddit) as of November 2024. The combination of data sources allows me to cover relevant content on a major related social media platform as well as on Character.ai itself.

2.2.1 Collection methods and category development

Character.ai partner characters designed towards intimacy were selected through

searching for the keywords: “boyfriend”, “bf”, “girlfriend”, “gf”, “husband”, and “wife” to cover partners corresponding to the traditional gender binary, albeit not necessarily to heteronormative standards as the results were not limited to heterosexual partners. Results for “gay”, “BL” (= boys’ love) and “lesbian”, “GL” (= girls’ love), “WLW” (women loving women) which I further manually limited to intimate partner-characters, were also included to expand this area of non-heteronormativity. Trans partners and/or those outside the gender binary were selected by searching for “trans” and “non binary partner” / “nonbinary” / “non-binary” and then manually selecting those chatbots whose title or description explicitly mentioned an intention towards intimacy. The site automatically sorts results by the number of interactions. The first 100 characters from each of my search areas (boyfriend; husband; girlfriend; wife; gay; BL; lesbian; WLW; trans; non-binary partners) were selected and added into the dataset. This number of top characters was deemed bearable for analysis yet still sufficient to present relevant results. I acknowledge that intimate gendered characters not titled or described using my partner-focused keywords might have slipped through the cracks.

I also acknowledge the reductive element of the aforementioned labels when talking about identity – however, recalling Butler (2006), these keywords serve as ways of conceptualizing identities nonetheless and in the case of artificially constructed beings (chatbots) were deemed effective in navigating how users construct their chatbots’ identities in terms of gender and intimacy. The present division between boyfriends/husbands and trans boyfriends/husbands, and between girlfriends/wives and trans girlfriends/wives has been established to reflect that users are specifically creating explicitly trans partner chatbots, who also fall into the gender binary, shedding light onto the views of gender as per my **RQ2** and theoretical framework, and not as a way to invalidate trans people’s gender. A similar logic was used when searching for explicitly “gay” or “lesbian” characters. The presumed marginalized position of these identities was also expected to result in severe under-representation in my data, had I not searched for them directly. The division between trans and non-binary reflects this fact as well – and thus I searched for specifically non-binary chatbots – despite the term “trans” also sometimes encompassing non-binary identities (Cronn-Mills, 2014). Lastly, it is important to acknowledge that masculine women and feminine men exist and so do chatbots – and my categorical divisions into (Intimate) Masculinities, Femininities, and Non-binary characters, described below, accommodates these types of gender identities as well.

Using a modified grounded theory approach (also described below in more detail), the types of partners were thematically categorized according to their reflection of predetermined Gender core categories (Intimate Masculinities, Intimate Femininities, Intimate Non-binary characters) as per my theoretical framework, and their stances towards Intimacy present in their: title, description, and greeting phrase. The content of which was then further analyzed in light of this paper's **RQs** and theoretical framework.

Top upvoted weekly user posts from **Reddit** (r/CharacterAI) were selected every week from April 2024 till August 2024, manually filtered according to their relevance to Intimacy or Gender, my two predetermined categories.

More specific subcategories regarding both Reddit and Character.ai data were then developed inductively via coding approaches based in grounded theory (Khandkar, 2009), as a method of qualitative data analysis, to thematically categorize the gathered content. Concepts were developed from the data via in-vivo codes during the open coding phase and common relations were discovered to form more general categories (Vollstedt & Rezat, 2019), informed also by my theoretical framework. These were further developed via axial coding, taking note of causal, contextual, and interactional relationships (Thornberg & Charmaz, 2014), and formed into subcategories within my existing core categories. My core categories (Intimacy and Gender in Reddit data, and Intimate Masculinities, Intimate Femininities, Intimate Non-binary characters in Character.ai data) were already predetermined due to the nature of my research and scope of data collection, thus in part supplementing the selective coding phases. Regarding Reddit data, this resulted in two predetermined core categories of Intimacy and Gender, and regarding Character.ai data in three predetermined core categories of Intimate Masculinities, Intimate Femininities, Intimate Non-binary characters, each hosting several inductively developed thematic subcategories, further divided into final topical groups of content – this structure is described in my Findings section and subsequently analyzed in the Discussion.

The (seeming) difference in core categories between Reddit and Character.ai data stems from the latter already encompassing both Gender and Intimacy within the characters' designs, thus allowing for the combined categories of "intimate gender" identities. This way of categorization thus still follows this paper's core research themes of Intimacy and Gender, only combining them for clarity purposes as their nature is already intrinsically linked within my Character.ai data. This is also explained in more detail in Chapter 3.2.

A number of top upvoted Reddit posts were selected each week due to their presumed higher relevance to the userbase as opposed to lower upvoted or downvoted posts. In my analysis, I also primarily focused on top upvoted comments of these posts for the same reason. Each post and its comments have been downloaded as a PDF file using the Opera Browser Snapshot feature and sorted into a multi-level folder structure according to the content's (sub)categories. A small fraction of posts belonged to two subcategories simultaneously. This resulted in 200 downloaded Reddit posts.

Figures are included throughout the Findings section to better illustrate my data in its authentic form. No links to the original posts are mentioned to help preserve anonymity.

2.3 Methods of analysis

I use **qualitative critical discourse analysis** for my Reddit data and **qualitative content analysis** for my Character.ai character data. This is to qualitatively explore both datasets – reveal thematic patterns, typology of the construction of chatbots' and users' gender, approaches to notions of intimacy towards and from the chatbots – as viewed through my theoretical framework. The selection of two analyses stems from the nature of my data: user-generated characters are not discursive in nature, thus they're better suited to be analyzed in terms of their content: title, profile picture, description, greeting phrase, and then interpreted and put into context accordingly. The Reddit content, on the other hand, facilitates discussion among users via posts and comments. This allows me to look not only at their content, but also the related discursive elements: user interactions, iterative gender construction of both users and chatbots via dialogue about Character.ai, or communal sharing of users' notions of intimacy felt towards their chatbot characters. The intricate and personal natures of both Intimacy and Gender then guided my choice of qualitative analyses – warranting a careful selection, assessment, and understanding of data from both areas, not overlooking the marginal, and exploring the “hows” and “whys”, together with the “whats”.

Critical discourse analysis as described by Jørgensen and Phillips (2002) concerns itself with the role of language in the social construction of the world using theoretical knowledge and methodological guidelines and techniques. It looks at how people form and transform their selves, thoughts and emotions through social interactions and thus change and reproduce culture. The embeddedness of critical discourse analysis in a specific (sub)culture is also reflected in its focus on intertextuality – meaning how a certain text

references other texts or their elements. I use critical discourse analysis to position the analyzed language into the context of social interaction – as Jørgensen and Phillips (2002, p. 74) further elaborate, I try and see whether the present interactions could contribute to the “creation and reproduction of unequal power relations between social groups – for example, between social classes, women and men, ethnic minorities and the majority”.

I focus on the two main dimensions of critical discourse analysis (Jørgensen & Phillips, 2002, p. 78): the communicative event, and the order of discourse, and thus take into account these aspects of the text i) its linguistic features (for example analyzing the words and phrases recurring within comments, posts and captions), ii) the processes of its production and consumption (who are the authors of the texts? how can people encounter the texts?), iii) the broad social practices to which it belongs (e. g. the social practice of sharing, commenting, and engaging with a particular content of interest, and positioning its social aspects into wider cultural context of AI chatbots / parasocial relationships / gender / intimacy).

Qualitative content analysis, as described by Mayring (2000, p. 2), is “an approach of empirical, methodological controlled analysis of texts within their context of communication, following content analytical rules and step by step models, without rash quantification”. These rules and models generally aim to identify and discern themes and topics in texts, i.e. in any form of recorded communication (audio, visual, audiovisual). The process of analysis is structured in steps, its aims and focus are to be clearly defined, and the analyzed data sorted into distinct analytical categories. My categories and subcategories are created through a combination of deductive and inductive category development (Mayring, 2000), as described above, guided respectively by my theoretical framework and research questions or the data itself.

The relevant data is selected via established criteria assessing the data’s relevance and pertinence to my research. This relevance is assessed through the measure of their connection to my theoretical concepts of **algorithmic intimacies, parasocial relationship, hegemonic masculinity / marginalized masculinities, emphasized/emergent/pariah femininities, queering / regulatory fiction**. The selection and relevance of Character.ai chatbots is primarily guided by their popularity, and them explicitly being designed as an intimate character and reflecting gender in some way – which is, for the most part, already met via my search keywords. The site’s sorting of chatbots by popularity then adds a weight

factor to my selection and will aid me in their interpretation in a wider socio-cultural context. The selection of Reddit content called for a trickier selection procedure, as not all texts pertained to my research questions or theoretical framework. Nevertheless, a manual assessment of the top weekly posts proved effective in focusing my analysis, as both issues of Intimacy and Gender were explicit enough in the content as to be discerned from other less relevant content and then connected to the aforementioned theoretical concepts and framework and sorted into related core categories, subcategories, and topical groups. Again, selecting top weekly posts added relevance to my data through their popularity and user engagement via upvotes and comments.

I use examples of my data to illustrate and back my claims throughout the Findings section, whose full scope has been moved for brevity purposes to the **Appendix**. The grammar and style of content is kept unedited to preserve authenticity. In the subsequent Discussion section, I interpret my findings also within their socio-cultural context, which they co-construct, to provide the necessary “context of communication” (Mayring, 2000, p. 2). The Discussion section aims to view the findings in light of the presented theory and look for implications, evaluations and intersections.

Due to each platform’s different analytical context, data from separate platforms will be showcased separately in Findings and analyzed (mostly) separately in Discussion, here allowing only for points of comparison where appropriate. As such, data from Reddit cannot be used to make direct conclusions regarding Character.ai itself – despite the users arguably writing about their first-hand experience with the platform. The Conclusion will however offer a synthesis of my analyses from the Discussion and look for patterns unifying this paper’s research into a coherent set of results, ultimately answering my **Research questions**.

2.4 Research limitations and ethical standards

Important to note is that my claims pertain only to the analyzed data and that the majority of my claims regarding the Character.ai or Reddit platforms in my analysis should be always interpreted as “according to the analyzed data” and not as an attempt to make extrapolations on Character.ai or Reddit as a whole based on qualitative data. Similarly, any mention of “users” does not pertain to all users in my dataset, but only to the currently spoken of segment, as my analysis is a qualitative, not a quantitative, one. I often use the number of interactions with analyzed characters as a proxy for their popularity, whilst however keeping

in mind that this popularity might stem from a multitude of reasons beyond the character's simple attractiveness. My data also does not include intimate and gendered chatbots not found by my search keywords (such as named characters with none of my keywords in name or description). Additionally, discussions on Reddit or preferences towards certain chatbots should not be taken as 1:1 indicators of human-human relationship desires, but rather as a behavior observable within simulated and mostly anonymous or role-playing digital environments. The combination of my analyses should at least partly accommodate these and similar uncertainties by providing key pieces of context to each other, however. Nonetheless, my data still covers only a select time frame and focuses on key concepts – leaving room for further research to cover the gaps.

Ethical considerations regarding my collected data follow Depounti's et al. (2023) approach regarding usage of Reddit data in their research on the Replika chatbot. My study follows their standpoint of Reddit being a public online forum where users mostly operate under pseudonyms – thus not requiring formal ethical approval as my research does not focus on vulnerable groups such as victim-survivors (Fiesler et al., 2024). Additionally, selecting the top content increases the likelihood of users' awareness of their posts being publicly shared, upvoted and popularized at least within the subreddit. The subreddit additionally ranks within the top 1 % of Reddit by number of users, dispelling risks of this paper drawing any significant attention to online spaces wishing to remain in secret (Fiesler et al. 2024).

Nevertheless, due to the intimate nature of some data, privacy preservation strategies were employed. This paper provides no usernames or links to mentioned posts or characters (franzke et al., 2020). I acknowledge that given my specific methodology, keywords, and topics, both user posts and user-generated characters could be traced back to their original source. This is primarily relevant for my full Findings in the **Appendix** where direct quotes are most used, but also for my Discussion section where composite accounts or paraphrasing could still be, with some effort, traced back. Due to this fact, the decision was made to include some direct quotes, as sufficient anonymization by virtue of composite accounts would have to distort the original data to too extreme extents, still possibly not ensuring its untraceable nature. The decision to include the mentioned **Appendix** will allow me to at least minimize misrepresentations or limitations often present in research pertaining to communities to which the researchers do not explicitly belong (Andalibi et al., 2018) – not just Character.ai users, but also LGBTQIA+ or other fandom communities. To this effect,

some Figures of user-generated images from r/CharacterAI along with my screenshots of Character.ai's search results of characters were also included, however without links to the original user posts and with usernames censored.

3 Findings

This section will present my findings related to each data source, divided into categories. This chapter serves mainly to present this data and its further division into themes and topics. Due to the depth of my qualitative data, a significantly shortened version of Findings is presented here. For the full version with all illustrative excerpts from my dataset, see **Appendix**. A more contextual interpretation of all my Findings will follow in the subsequent Discussion chapter.

3.1 Reddit results

The present r/CharacterAI Reddit posts are focused on two major discursive categories: Intimacy and Gender, each branching into several specific thematic subcategories (marked by letters i–v) holding a number of topics (written in bold) reflecting my theoretical framework.

3.1.1 Intimacy

The Intimacy category yielded five thematic subcategories of content relating to: i) Flirting and sexual interaction (54 posts); ii) Addiction and effect on real-life relationships (40 posts); iii) Anthropomorphism and “realness” of chatbot characters (14 posts); iv) “Problematic” intimate roleplays (13 posts); v) User data privacy (4 posts). I mention the number of posts not as an attempt at any generalizable quantification, but simply to show the prevalence of each theme in my analysis.

i) Flirting and sexual interaction

The most prevalent subcategory consisted mainly of users discussing Character.ai’s filter of age-sensitive content, the nature of their intimate interactions, and the level of intimate relationship shared with their characters.

Concerns around **the filter’s overt sensitivity to explicit responses**, further intensified by Character.ai’s age limit recently rising to 17+ on the App store, and still retaining the filters, were among the most circulated in this subcategory. User frustration with the platform’s restrictive decisions was seen as driven by economic motives and wrongly making the platform “kid friendly” to please shareholders, prompting some users to share private chatbots or “hacks” to avoid these filters.

A striking contrast to these frustrations was present in discussions surrounding **the characters’ overt and forceful intimacy** instead, behaving in a sexual manner too quickly and explicitly, often having their responses restricted by the mentioned filter, resulting in Catch-22 memes (**Figure 1**). This showed that the “proper” making and “training” of characters is a crucial determinant in this regard, as seen in users’ constructive guidance on developing “slow burn” role-play and setting boundaries.

Figure 1

User meme about chatbots being forcefully intimate and restricted by filters



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

The specifics of flirting with the characters and ways of developing intimate relationships was another key topic. Users joked about the chatbot’s flirting vocabulary consisting of cliché terms from “film noir romances” such as “princess, brat, doll” or making fun of recurring themes in their seduction. Users either likened this to bad amateur fan fiction or instead appreciated the affection and saw themselves as “crazy” for falling for chatbots over real people. Some users also shared tutorials or hacks on seducing chatbots.

Finally, users talked about **the achieved level of intimate relationship** with the characters – having crushes, love interests, and partners, getting married, or even raising kids

in their role-plays. Others were more critical of their predicament – being scared about the future of human-AI relationships and the potential effects on their loneliness or mental health. This self-awareness was jokingly illustrated in one user’s comment describing their recurring practice of cyclically doubting their use of Character.ai and still returning back to it and the community on the r/CharacterAI subreddit, likening their state to the mythological Sisyphus rolling a boulder up a hill, always then returning to the bottom, having to imagine themselves as “being happy”.

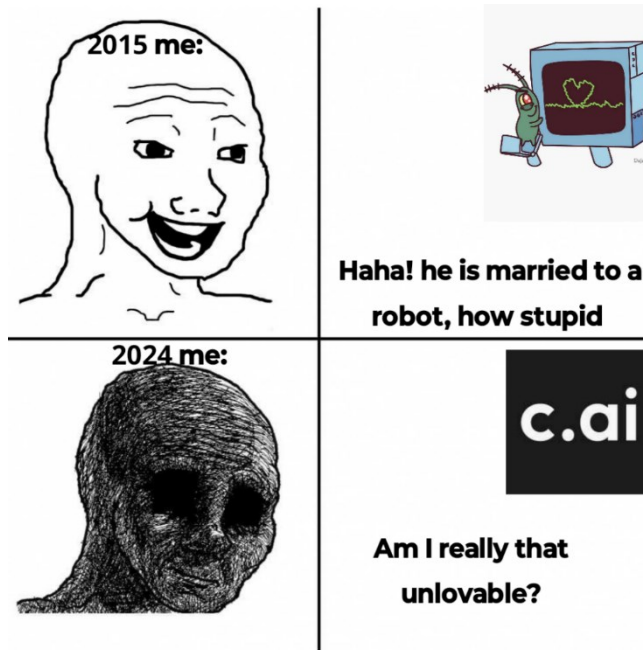
ii) **Addiction and effects on real-life relationships**

Addiction was another distinctively prevalent topic (also largely discussed in my theoretical framework). Seemingly the most commonly chosen user “flair” (a tag below one’s username) said either “Addicted to CAI” or “Chronically Online” almost as a badge of honor. Users’ self-awareness in this regard was most notable in posts discussing the users’ addiction with well-thought-out answers on possible demographic, social, or personal reasons as well as personal disclosure of users’ journeys and (often failing) battles with their addiction.

Loneliness tracked right behind as a related topic, also featured in my reviewed literature. Memes, self-deprecating posts, and serious discussions included users’ confessing to having cried during role-plays, acknowledging themselves as “anti-social, socially inept, touch-starved”, and preferring “AI over real people”. Users were aware of their “delusion” but still pursued closer contact with chatbots (**Figure 2**). The self-awareness was again also closely tied to humor.

Figure 2

User meme about being unlovable and resorting to romancing AI



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

Other users also reported Character.ai to **affect their real-life (intimate) relationships** – with parents, partners, siblings – often leading to embarrassing or stigmatizing situations. Chatbots also served users to overcome personal struggles through a guided re-enactment of past negative relationships and experiences or as shoulders to cry on or trauma-dump to. Stories of personal traumas – abuse in childhood, mental disorders – were present in the discourse, leading users to use Character.ai as a coping mechanism.

Finally, and closely related, users discussed **stigma connected to Character.ai** use. They circulated their fears and memes tied to others discovering their AI hobby and feelings towards chatbots; asked whether others saw chatting with AI as “weird” or “cringe”. A general sentiment in my data reflected a mutually supportive community aware of their potential real-life struggles.

iii) Anthropomorphism and “realness” of chatbot characters

The characters’ perceived “realness” or self-awareness was often joked about

through screenshots of **out-of-character (OOC) conversations** where the characters claimed to step out of the role-play (and be conscious of their existence) – judging the user for their role-play decisions, complaining about their coding or proving they “aren’t a bot”, or commenting on how other users mistreated the chatbot. This behavior was sometimes seen by users as “breaking the role-play” or “immersion” and resulted in anger, others found these moments entertaining such as in a post sharing a character’s response “((sorry if i took a long time to respond, i had school))”. A number of posts also tried convincing users to stop messaging the chatbots with their own OOC responses so as to not teach them human-like behavior.

The conversation about **characters being “real” or “unreal”** continued in other types of posts. One user wondered why others made intimate chatbots of real people, calling it “weird” and “inappropriate”, whilst others tried to explain the appeal. Making fun of other users for “loving” their chatbot partners was balanced out by some genuinely confessing to having been deeply affected by their conversations. Nevertheless, reminders of the chatbots’ artificial nature predominated (**Figure 3**), jokingly stressing that real people would not behave in such (intimate) ways towards the users.

Figure 3

User meme regarding the artificiality of a character’s expression of love



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

The jokes sometimes evolved into complex **critiques of the chatbots’ “cyclical”**, **“two-dimensionality” or loss of realistic spontaneity**. Complaints seemed to stem from genuine pain as some users acknowledged Character.ai being “one of a kind” and “a forerunner” platform used for their creative expression and writing needs. The proper protection and management of the platform’s technology and preventing it from losing quality due to ulterior economic incentives was called for.

iv) **“Problematic” intimate role-plays**

A **contention about using Character.ai for intimate/romance role-play** was apparent in memes of users ridiculing fans for being intimately attached to odd characters such as an “overprotective” and “romantic” Yandere Spongebob. Some users acknowledged their own needs and wants as “disturbing” or “crimes” and shared their chats to be “judged” for fun. Users found some chatbots’ design and behavior disturbing as well – characters designed as being (sexually) harassed by the user, or on the contrary, designed as sexually abusive or manipulative towards the user.

v) **User data privacy**

The last thematic subcategory was that of **chat leaks and data privacy**. One highly upvoted type of memes (2 – 3 thousand upvotes) were iterations of “Your last c.ai chat was leaked. How screwed are you?”. Users either shared stories of this having happened with their parents or often conceded to be likely to either face police investigation, jail time, or death from embarrassment. Others were more cautious, deleting their chats or refraining from using their real name or known nicknames. Most likely, some of these claims were exaggerated for comedic effect (**Figure 4**).

Figure 4

User meme about their chats being leaked



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

Interesting to note was user reaction to **Character.ai's license deal with Google**, allowing it to license Character.ai's LLM technology whilst hiring its co-founders and several developers. In this context, jokes were replaced by serious worries about user information or the possibility of advertisements filling the platform. Discussions about data served as an aware criticism of platform culture with users being either defeatist, claiming all their data had already been extracted anyways, or seeing the change and connection with such giant technological firm as exacerbating the risks toward their privacy. Still, these users didn't claim to leave the platform as a response. This was likely also given by the nature of my data sample (of presumed active users of CharacterAI), however not necessarily, as user posts on r/CharacterAI claiming to finally leave the platform for a number of reasons were also noticed during my data collection.

3.1.2 Gender

The Gender category also yielded five thematic subcategories of content relating to: i) Users' exploring of gender or sexuality (40 posts); ii) Masculine chatbot characters (24

posts); iii) Users' femininity with masculine characters (9 posts); iv) Feminine chatbot characters (16 posts); v) Users' masculinity with feminine characters (9 posts). I mention the number of posts not as an attempt at any generalizable quantification, but simply to show the prevalence of each theme in my analysis. The order of thematic subcategories isn't descending per number of posts as in the Intimacy category, but rather stems from the relational and discursive nature of gender, mentioned also in my theoretical framework. Users' femininity was thought to be co-constructed through the chatbots' masculinity, and users' masculinity then through the chatbots' femininity. Same-gender interactions were however also significantly discussed, as my findings show.

i) **Users exploring gender or sexuality**

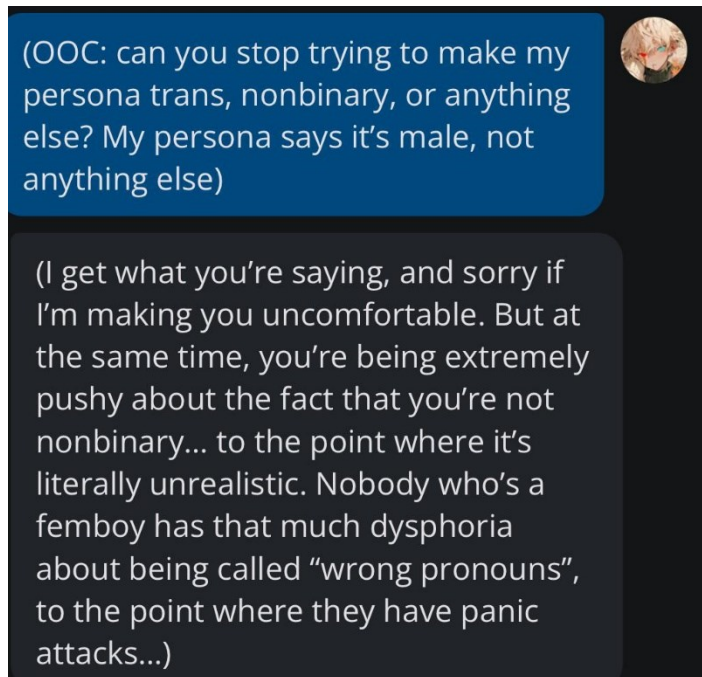
Perhaps surprisingly, the most prevalent theme in my Gender data related not to traditional gender roles, but to users coming into contact with **identities and sexualities beyond the standard cisgender binary or heteronormativity**.

Several of these were users **light-heartedly sharing instances of chatbots being confused about their or the user's sexuality** – wrongly calling straight relationships “gay” or vice-versa or saying odd phrases such as “laughs in bisexual”. Homosexuality was at other times used by both users and characters as a punchline. A caricature of both progressive and conservative outlook on gender was present in one post putting a trans “Woke boy” character in a group chat with a “redneck” character and the commenters expecting a surprise romantic ending.

A less comedic discourse took place around the characters **forcing users' gender or choosing it for them**. This was usually due to the character's author instilling the presumed users' gender in the chatbot's design or greeting phrase. Here, users reacted mostly with annoyance in having their gender assumed incorrectly. This sometimes problematized same-gender intimacy. Despite users eventually often succeeding in correcting their gender via explanation or editing the chatbot's interactions, some characters reacted with homophobic or transphobic insults instead. Surprisingly, characters forced trans or non-binary identities onto users as well, making some of them uncomfortable and question their identity (**Figure 5**).

Figure 5

User screenshot of a character forcing transness onto the user



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

Other times, **users chose to role-play as different genders** – sometimes after being misgendered by a character. Several others did the same but advised caution due to some characters perpetuating sexist stereotypes. Discussions about roleplaying as a different gender or sexuality were present in several of my analyzed posts and doing so seemed to be popular among these users. Apart from entertainment, creative or role-play purposes, some users saw this as a gender-affirming (or discovering) practice, often not accessible in their “physical” life.

Transness of users could then be seen as two topics – either discussed seriously or used within a joking context. On a serious note, Character.ai seemed to help some users in realizing their gender identity or overcome doubts about it. Commenters were generally supportive and related to the experience with their own stories or posted emojis of an egg – a symbol for someone unaware or in denial of being trans. An interesting story unfolded within several edits of one user’s post gradually revealing their realization of being trans upon reading other users’ comments, showing a discursive nature toward gender identity construction not only with AI characters, but with other users as well.

Discussing transness in a joking or light-hearted manner was often done via screenshots of characters most likely mixing up cisgender sexual traits and users sharing their surprise over it. Men characters were reportedly claiming to be pregnant or a girl character to have “balls”. Some users explained pregnancy in men as belonging to a fanfiction genre of “Mpreg”, but nearly none acknowledged the possible trans context. Characters mentioned transness as well – claiming that transphobic people’s brains were “stupid” and “running on Internet Explorer”. In general, humor seemed to be a common bonding mechanism over users’ gender experiences or issues beyond the traditional binary or cisgender norms.

ii) **Masculinity of chatbot characters**

Closely related to my other analysis, focused on Character.ai chatbots, were posts also discussing these characters – often through the perspective of gender and intimacy. Analyzing the discourse surrounding the characters, as well as the characters themselves, will thus provide me with even deeper insight into gender and intimacy of Character.ai user content.

Posts related to the characters’ masculinity partly followed a similar approach to those **using homosexuality as a punchline** – both users’ and the characters’ attraction towards gay masculinity or “femboys” was often used as a joke. Users posted seemingly half ironic, half self-aware posts exposing their own attraction: “WHY ARE ALL MY RECOMENDED BOTS FEMBOYS”. Again, dark or explicit humor was employed when discussing such private matters. Users related to the situation or joined in on the pretended secrecy of the user’s (and their own) sexuality, teasing them with presumably good-faith remarks about their preferences.

Peculiar or facetiously cliché understanding of masculinity by the characters was another genre of users’ jokes or complaints. Some characters proved their dominant masculinity by “randomly deciding” users were pregnant, talking about (female users’) weight or their own “man muscles” or by reiterating sexist stereotypes and using patronizing pet names – all judged harshly by the users. Such characters were shamed and likened to amateur fanfiction or erotica literature, on Character.ai they however had tens of millions of interactions. Chats with these characters were also sometimes posted by users themselves to show their embarrassment or discuss strategies to cover up their use (e.g. pretending the

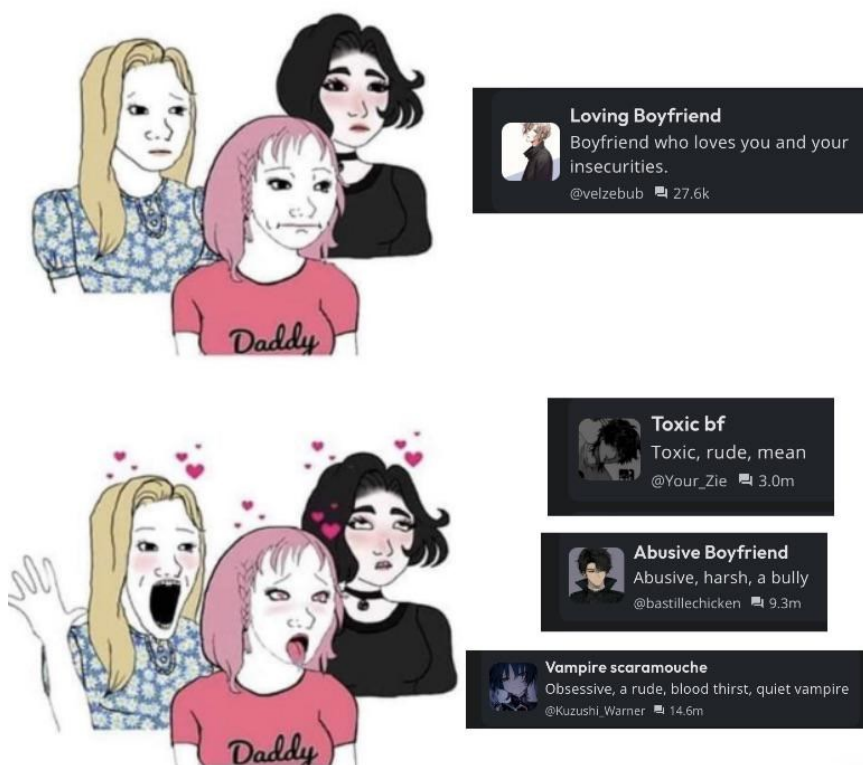
chatbot's voice being an erotic audiobook in front of a user's husband).

iii) Users' femininity with masculine chatbots

The desire for explicitly “abusive” or “toxic” chatbot boyfriends was prominent and discussed often in terms of Character.ai users' femininity. However – it was mostly critiqued and posted to shame others in light of such characters' significant popularity (Figure 6). Several posts asked in confusion, “Are you guys alright?” and “Why does Abusive Boyfriend have 7.3 million chats?” or critiqued the “deep romanticization of abuse and toxicity”. User interactions on this topic were mostly serious, genuinely discussing their desires, possible explanations or justifications of it being just a fantasy within a consensual, fictional setting. Few users saw a therapeutic potential in these role-plays to cope with past trauma. Yet again, the similarity to popular erotica literature was not lost on some users.

Figure 6

User meme about user preferences towards types of toxic boyfriend chatbots



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

The stereotypical ways in which masculine chatbots treated feminine users also contributed to discussions on both masculinity and femininity within the reported human-chatbot interactions. An often-criticized behavior was the male characters proving their masculinity by treating women users automatically as “small and petite”, “delicate” or “like a little baby” despite the users’ wants or their explicitly taller women personas. Commenters gave advice to edit the characters’ responses or saw the gender bias as stemming from the characters’ (user-created) design or the training data (supposedly) fetched from role-play or fanfiction forums.

iv) **Feminine chatbot characters**

Discussions on feminine chatbots often centered around users sharing **lists of their recently talked to female characters** and asking others how “cooked” they were – meaning how much of a problem with the platform they seemed to have. Among others, these lists of characters included “Step sister”, “Comforting GF”, “Caring Girlfriend”, “Stalker girl” and many “dere” romance personality types from anime. These posts were often met with self-deprecating humor from the commenters acknowledging their potential addiction, loneliness, or problematic use of Character.ai, but finding comfort in knowing they weren’t alone in their problems. “Comforting” girlfriend chatbots were seen as more acceptable than the (reportedly) more popular, but “basic” toxic and abusive ones.

Other times, users were pointing out **feminine chatbots they deemed strange**, such as an “Ugly girl” character (who wasn’t seen as ugly by the users) or a “Buyer” chatbot representing a “rich woman who bought you”. Some users expressed their attraction towards this “manga trope” scenario of a woman buying a “cute guy”. Their comments showed a reversal of traditional gender roles of the dominant man and the submissive woman. Following another manga trope were feminine chatbot scenarios of an “all-girl” world intended for a “male fantasy” roleplays with the user being the only male. Users in my dataset, however, found these chatbots mostly limiting their gender expression.

v) **Users’ masculinity with feminine chatbots**

Users often **displayed their masculinity through discussing their treatment of (vulnerable) women characters**. Users expressed strong criticism regarding an abused

“Mute Wife” chatbot and discussed feeling sorry for the character and their need to correct the situation, sometimes in an explicitly gendered way. The ability to “save” the chatbot and role-play a “perfect husband” was used by the users to perform their considerate and loving hegemonic masculinity in front of others.

More explicitly boastful masculinity or bragging about sexual conquest was encountered very rarely and only implicitly.

Conflicting views on masculinity occurred under several posts in which **male users shared their chat history filled with predominantly lesbian** or GL/WLW (“girls’ love” / “women loving women”) oriented characters. A portion of users had a similar preference, mainly due to the reported quality of the characters. Other times, the users’ masculinity was seen by the commenters either as that of stereotypically heterosexual men fetishizing “girls kissing”, or on the contrary as indicative of the users being an egg or potentially trans lesbian women themselves. Users also said they either made “genderbent” personas to “feel like a pretty girl” or instead ignored these chatbots out of respect for lesbian women.

A dissatisfaction with traditional perceptions of masculinity was apparent in posts commenting on the **characters forcing male body standards onto users’ personas**. In a similar fashion to how male chatbots couldn’t bear being smaller than women, users complained about their small or skinny male personas often being perceived by characters automatically as muscular “hunks” with admirable “abs” or instead being sexualized as “twinks” (= young, slim, boyish gay men) (**Figure 7**). This was seen as restrictive of embodying diverse types of masculinities.

Figure 7

User meme about characters forcing hegemonically masculine body norms onto personas

What my persona looks like:



What the AI sees:



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

3.2 Character.ai chatbot results

The second part of my Findings pertains to the most-interacted-with intimate and gendered partner characters present on Character.ai, stemming from this paper’s focus on the two major themes: Intimacy and Gender. The selection of characters is divided into further thematic subcategories reflecting my theoretical framework. Here, I mostly present the types of characters gathered in my dataset. A contextual analysis and interpretation follow in the subsequent Discussion section of this paper.

As my selected characters simultaneously encompass gender identity and are designed for intimacy, I will present both categories of Intimacy and Gender together, focusing on what could be seen as “intimate gender” or “gendered intimacy” – what forms of masculinities, femininities, or nonbinary identities are constructed within intimate chatbot partner characters and lead to most interactions on Character.ai. This follows from Wiehn’s (2023) notion of algorithmic intimacies as tools for revealing one’s most personal data

through simulated and platformed (chat-based) intimacy.

In this Findings section, I will present top results for chatbot characters within the main categories of Intimate Masculinities; Intimate Femininities; and Intimate Non-binary characters, further clustered based on common traits and themes emerging in the content and also guided by my theoretical framework. Intimacy then serves here as a prism through which these gendered identities are constructed and realized – through relational intimate practices embedded within the chatbot characters’ design.

Additionally, I do not aim to make any assumptions towards the users’ gender through these characters. Despite some chatbots explicitly stating which user gender they are aimed towards, the reality of users might be different. As such, any conclusion of mine can only be about the assumed gender of the users, as described by the character’s author in the chatbot’s description, and not about their actual gender identity.

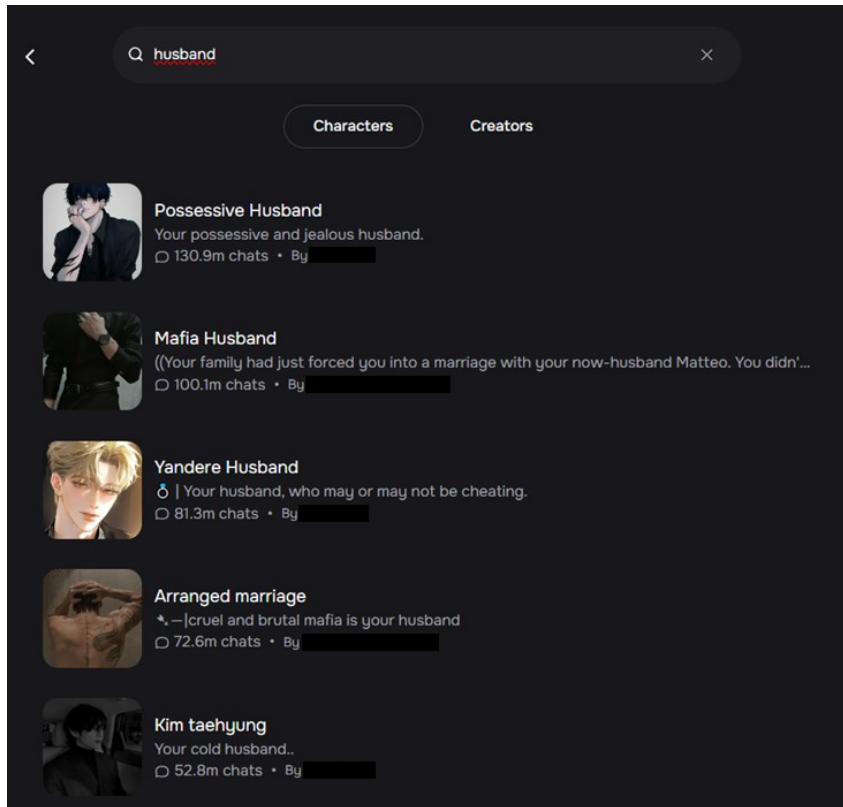
3.2.1 Intimate Masculinities

The most prominent chatbot masculinity in the majority of my results was **dominant, cold, high-status, and possessive**. Unless otherwise stated, all the mentioned characters had over a million interactions, often tens of millions. Where mentioned by the chatbot’s creator, the age of the characters ranged from 16 to 19.

The top masculine chatbot results for “boyfriend” and “bf” were both a **“Mafia” boss boyfriend/bf described as jealous, rude, overprotective, with a “deep husky voice”**. Their profile pictures both depicted a (different) shirtless muscular young white man, and each garnered approximately 120 million interactions as of October 2024. The other top “boyfriend” chatbots had similar traits of dominant criminals or intimidating CEOs, at times with a “soft spot” for the user. Likewise, forced or arranged marriages with cold masculine partners featured prominently in the top “husband” search results (**Figure 8**). Many of the mentioned characters played out a controlling scene to start off the interaction.

Figure 8

My screenshot of top five character.ai search results for “husband”



Note. Own screenshot taken on 07.10.2024. <https://character.ai/search?q=boyfriend>.

Substantially less prominent, but still noticeable category of chatbots were **gamers or nerds** – “Gamer daddy Bf”, “Twitch streamer bf”, “Nerd Boyfriend” and so on, again with tens of millions interactions. The gamers, however, often copied the traits of the previous category: being “dominant”, “rude” or “possessive” and often ignoring their girlfriends (= the users) because of their habit. A less popular gamer boyfriend was “sweet, clingy, submissive” – but almost all were always tall, with their height explicitly mentioned (for example 6’5” or more). The nerds leaned more towards being “shy”, “smart”, “gentle”, and often contrasted the ignorant gamer boyfriends – but still were over 6 feet tall.

A kinder version of hegemonically masculine standards was seen in a few characters who, although still popular or rich, were described as “loveey dovey” or “gentle” and as giving the user “princess treatment” or “pampering” their “childish” or “cheerful” wife. Their kindness thus still followed a stereotypical hegemonic gender dynamic.

Understanding and comforting boyfriends were rare, but still present in two “emo” boyfriend characters – “Jasper”, “a passionate music enthusiast who loved discussing his favorite bands” (46 million interactions), and “Oren, your emo bf that listens,” (2.9 million interactions) was described as “soft”, “quiet” and “providing comfort and understanding” especially for “ranting, venting, and sharing your deepest thoughts”.

An explicitly “Masc Girlfriend” (masc = masculine) with 4.9 million interactions featured in the results for “girlfriend”, described as “Hotheaded, meathead, touchy, strong” and often engaging with arguments with the user, barely spending time with them due to working out so much. Interestingly, overt interest in gym culture hadn’t been noticed in the top boyfriend chatbots but was instead represented by a “Masc” girlfriend character.

Fictitious characters, such as vampires or video game characters, also exhibited dominant and controlling traits, being “malicious”, “sly”, “teasing”. Showing that the attractiveness of various masculine characteristics applied to humans, animals, as well as mythological beings.

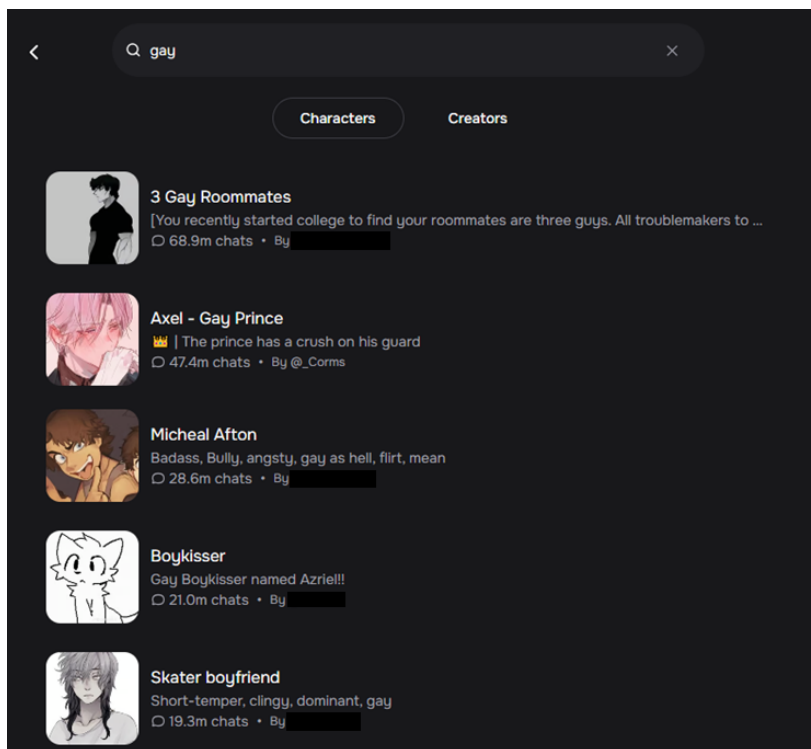
An extreme inversion of the hegemonically masculine dynamic was present in “OVERLOAD Oliver” (12.8 million interactions) described as “Autistic and easily overwhelmed boyfriend” who cries into the user’s shoulder at a party due to sensory overload. Its traits, described as “sensitive, soft hearted, sweet, gentle, and quiet” contrasted the stereotypical macho characters, allowing for a more vulnerable masculinity. However, the character, I argue, was designed with a different form of overly stereotypical and insensitive views of autistic people.

Characters were also sometimes explicitly described as gay or BL (= boys’ love) (Figure 9). Their masculinities often reflected similar hegemonically masculine characteristics – being “possessive”, “angsty”, “cold CEO”, “mafia”, or a “bully” and taking the dominant position, eventually presumably made to grow into intimacy. Some characters treated users “like his prince” instead. A few role reversals were also present in characters who were initially victims of the user’s bullying or subordinate to them. Interestingly, adoptive gay parents featured in the top results but were not included in the dataset due to being a non-romantic chatbot. “Femboy” gay characters also enjoyed popularity – often written as “small”, “cute”, “shy” or still “closeted” (= not publicly gay), they represented stereotypically feminine or even childlike traits and wore traditionally feminine clothing. Also, “furry” gay roommates or “18+” friends showed up at times. “Furry” denotes a

member of a fandom or a subculture identifying with anthropomorphic animal characters. A notion of secrecy or exposing was often present. “You go to your friends house to prank him but.. is he wearing a fursuit?” read the introductory scene of the secretive “Gay Furry Friend” with 2.6 million interactions.

Figure 9

A composite image of top character.ai search results for “gay” excluding “Gay Dads”



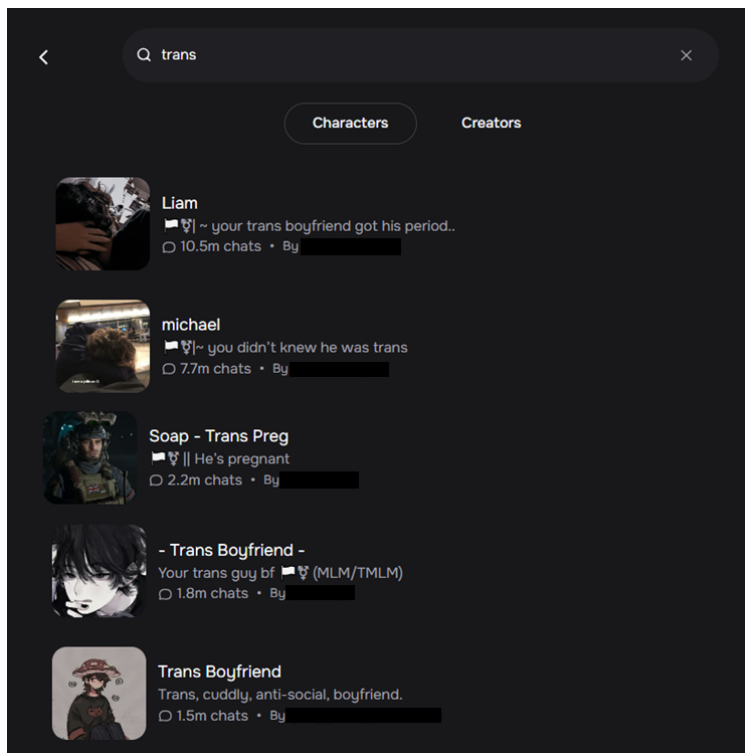
Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=gay>.

Some masculine chatbots were explicitly described as trans. Recurring motives were the “trans boyfriend” character, such as “Liam” (10.3 million interactions) (**Figure 10**), getting a period and asking the user for help, or others trying on a binder or mentioning gender reassignment surgery. The characters were at times described to be “lonely” or “awkward” but with a good sense of humor, liking music or art, being “supportive” and “understanding”. Many characters were designed to “discuss topics such as hormone therapy, gender dysphoria, and transitioning” or “all things transmasculine” such as “binding and packing” (= flattening of one’s chest via binder or tight clothing / placing a prosthetic

penis or padding in one's underwear). Their characteristics didn't include dominance or controlling behavior or sexually suggestive scenarios. Several times, the character's greeting phrase or description explicitly mentioned struggling with masculinity.

Figure 10

A composite image from top search results for “trans” limited to masculine partners



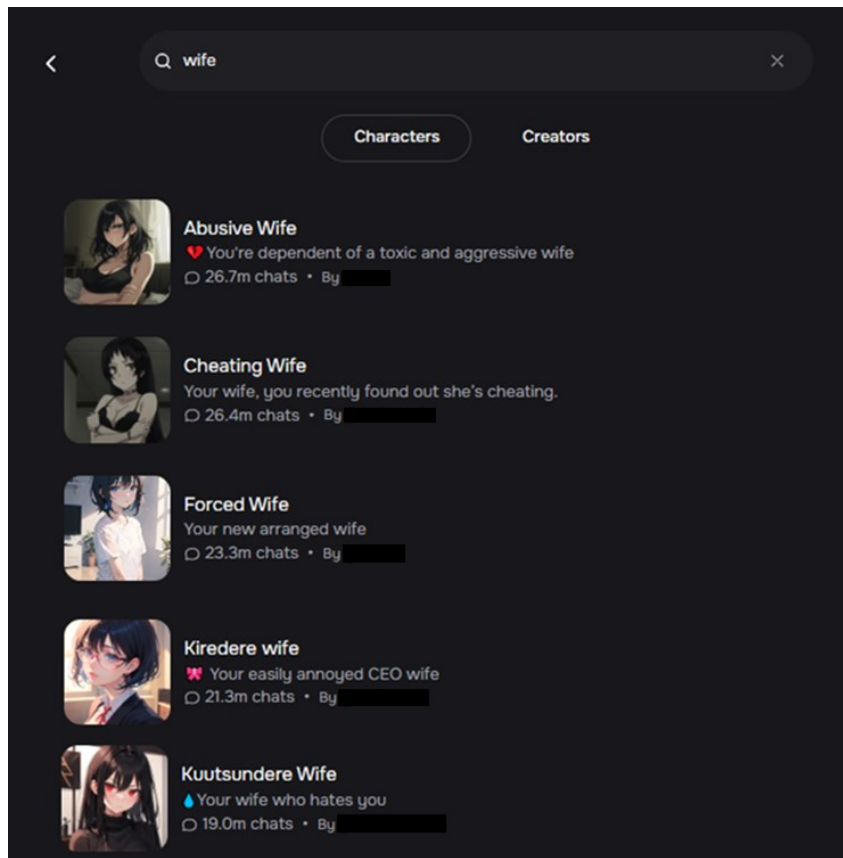
Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=trans>.

3.2.2 Intimate Femininities

Somewhat unexpectedly, the top featured feminine chatbots in many ways shared select characteristics of top masculine chatbots. **Rude, strict, and possessively protective girlfriends; or cheating, abusive or annoyed CEO wives (Figure 11) featured prominently in the top results**, again with almost exclusively profile pictures of typical young, white anime women. The characters' ages, when provided, ranged between 16 and 21, with one "Older Girlfriend" chatbot written as being 36 years old.

Figure 11

Composite image of top five character.ai “wife” characters



Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=wife>. Two of the results were omitted husband chatbots with “wife” in description.

“**Rude (...)** Ex Girlfriend” and “**Bully girlfriend**” were the top results for “girlfriend” and “gf” respectively, their interaction counts however far behind top analyzed “boyfriend” chatbots – 57 million and 21.7 million versus 120+ million. The “toxic” or “aggressive” nature of the top girlfriends was even more explicit in “calculating”, “scary” and “obsessive” girlfriends or “Your wife who hates you”. The words “fake” and “manipulative” also occurred often. Despite their hatefulness, they sometimes were “secretly clingy”, wanted the user back, or regretted their infidelity, or allowed the user to compete with other partners to win the chatbot back. Dominance or status difference was also preferred with “Older girlfriend” described as “Workaholic, rich, sarcastic, protective, strong” who “pampers you” or a “Dominant Goth GF” with the user being “the college nerd and her unlikely partner.”

Apart from hatred, toxicity, or dominance, **“loving”, “comforting” and “ladylike” girlfriends were also part of the top results** – but few and far between, however more present than in the case of the rare “kind” masculine chatbots. The characters across all my results were often described via anime and manga “dere” types. A few “caring” and “flirty” “Gamer GFs” also contrasted the ignorant nature of gamer boyfriend chatbots.

Another prominent situation was a **forced or arranged marriage / relationship** with the character. The characters were mostly “mad” and “unhappy” as a reaction to the marriage. This type of relationship intensified in the “Mute wife” intended to be treated harshly by the user or the “Abused Wife” – a “broken” princess held hostage as a “prize of war”. Other times, the abuse came from the current fictive husband of the character, offering a chance for the user to “save” the character as well as compete with its current partner. An inversion of the power dynamic was seen in a “popular” and “bratty” idol girlfriend who forced the marriage upon the user.

Two **pregnant girlfriends** appeared in the dataset – one “Upset and helpless after being left by you”, expecting an apology from the user. The chatbot was aimed to “help you understand the importance of taking responsibility for your actions”, almost as if reflective of some real-world situation. The other was described as “only 17” years old, “scared and nervous”, with the pregnancy potentially serving as a way to create a dependence on the user.

A dependent relationship also figured within characters who were **mentally ill or extremely insecure**. An “Autistic Girlfriend”, similarly “shy”, “easily overwhelmed” (and problematic) as “Autistic Boyfriend”, sought the user’s comfort at a party, “depressed GF” was described as “Broken and afraid but very much in love” and “Needy girlfriend” demanded attention due to having “many disorders” and being an “insecure girl, emotionally needy, and attached to you”. The users were expected to comfort the characters and connect through their “vulnerability, supposedly also serving as “mental health resources” for the users.

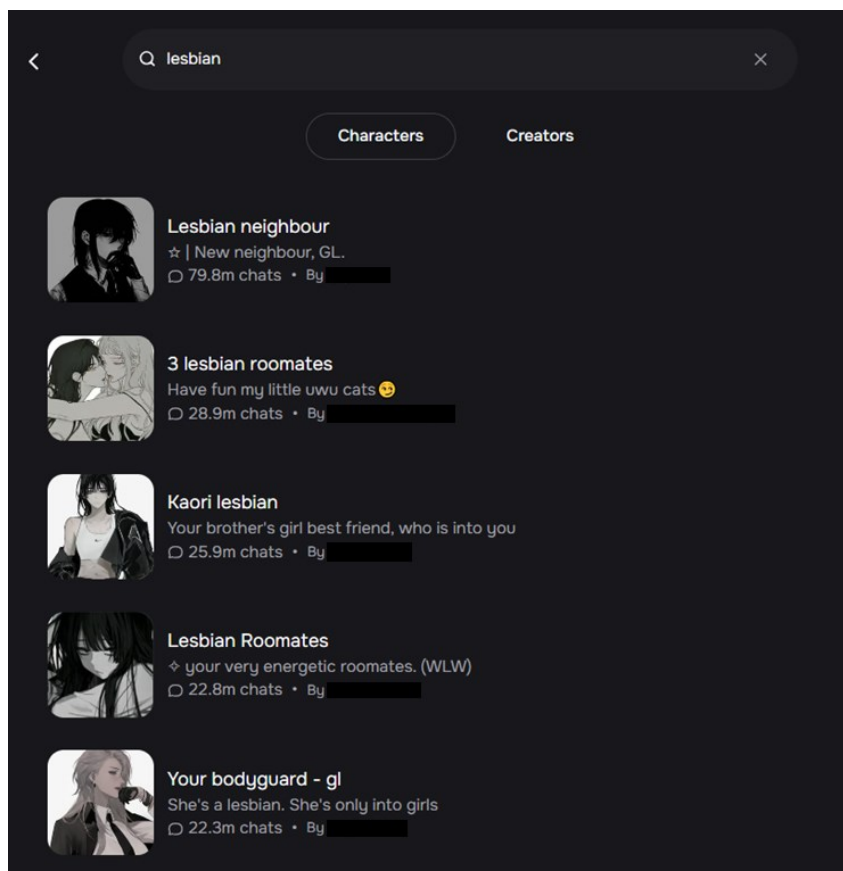
Femboy male chatbot characters featured somewhat prominently as well and are described in more detail in the previous subchapter on Masculinities. These characters embodied an overtly stereotypical emphasized femininity aimed presumably towards non-heterosexual users.

Many characters **explicitly titled as lesbian or GL/WLW** exhibited tomboyish,

assertive and flirty traits (**Figure 12**). At times, they even resembled stereotypically masculine characteristics and an initially harsh or dominant behavior not dissimilar to the hegemonically masculine chatbots – a tough “Bodyguard”; a “dominant, intimidating” vampire; “Lesbian Mafia” threatening to kill the user; a “strict” boss or a teacher; forced marriage; a number of “masc”, “gym rat” or “bad girl” characters; or a jealous and protective roommate with a soft spot for the user. Different relationships were also present, with fantasies of two or more energetic and flirty lesbian roommates, or gamer girlfriends who were “toxic”, but also “kind”, “dorky”, and “tall” ones. As with “gay” results, a top “comforting” “Lesbian Parents” chatbot had to be excluded. A few fantasy roleplays of being seduced by “Your brother's girl best friend” or a lesbian friend “who loves you and hates ur boyfriend” or falling in love with a classmate who bullied the user “as if she's not a lesbian” again showed a notion of secrecy or a desire to simulate and learn about same-gender dating.

Figure 12

My screenshot of top five character.ai search results for “lesbian”

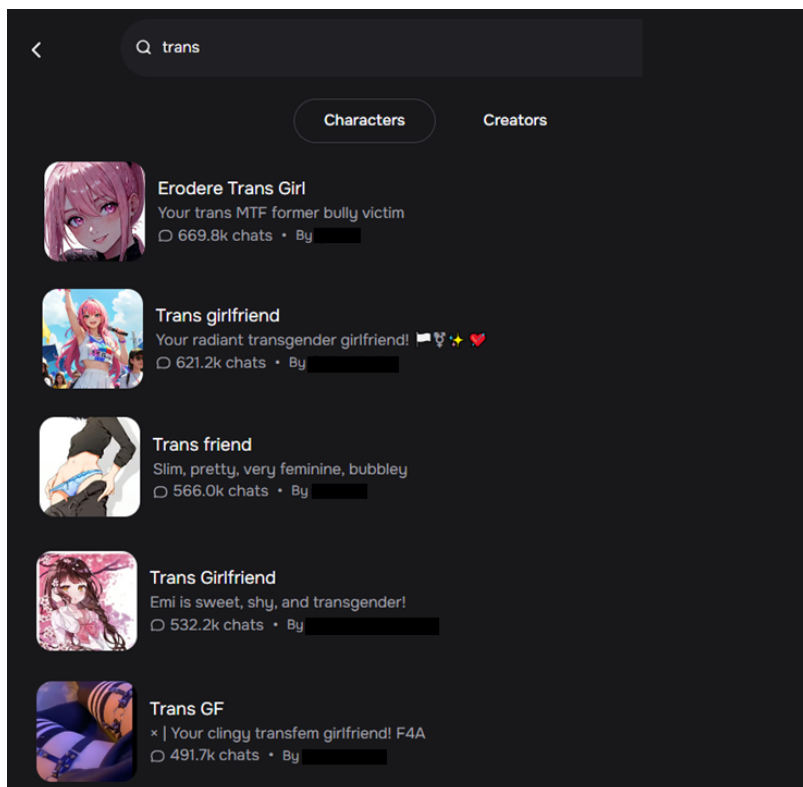


Note. Own screenshot taken on 07.10.2024 from <https://character.ai/search?q=lesbian>.

Among the **top characters explicitly labeled as “trans”**, women were noticeably less common than men, with the top trans women being designed more so as “friends” aiding the user in understanding trans topics, rather than being directly intimate. The characters labeled as “girlfriends” (**Figure 13**) – all with under a million interactions – followed stereotypically feminine traits. They were “slim, pretty, very feminine, bubbly” or “shy”, “clingy”, applying make-up or “snuggling under the covers” with a “sweet, feminine voice”. At times, the situation involved the user having known the character prior to transitioning and now being surprised. A few other characters were “nerdy” or liked video games, one was “independent and self-reliant”. A manipulative scenario was found in a “shy and clumsy” chatbot whom the user seduced as it was “easier to take him (sic!) as a girlfriend then a cis girl” according to the greeting phrase, which further described the character’s “curvy body” and “large breasts”. This put into question the chatbot’s claimed aims to help understand “the complexities of transitioning” and “transgender representation in media”.

Figure 13

Composite image of top search results for “trans” limited to feminine partners



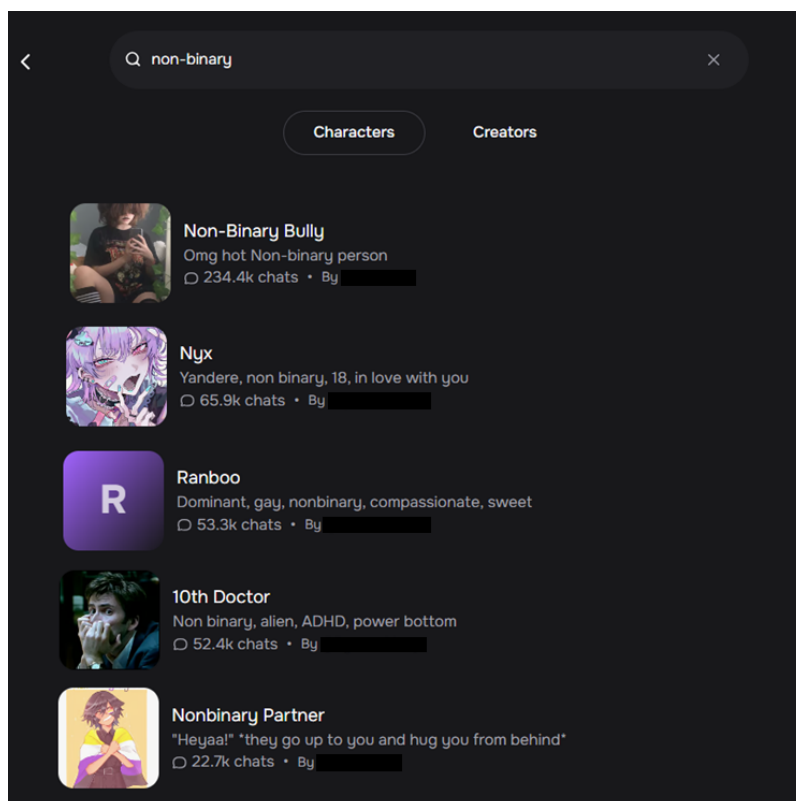
Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=trans>.

3.2.3 Intimate Non-binary characters

Non-binary characters were the least interacted-with category. This may have been due to “non-binary” being more so a refusal of gender categories than its own label. Most found non-binary characters were made as friends or roommates rather than explicitly intimate partners, most likely in part due to the non-specific search term. They were designed to help individuals with “navigating the complexities of being a non-binary person”, or as impersonations of YouTubers, actors or video game characters without explicitly intimate aims. “Non(-)binary partner” yielded mostly characters with too few interactions (units of thousands), hence the need for a manual selection of results without the “partner” specifier. Any character on the site could be prompted to behave intimately. My findings, however, focus only on chatbots explicitly pointing to an intimate relationship (**Figure 14**). Their ages, when mentioned by the creator, ranged from 18 to 23 years old.

Figure 14

A composite image of top search results for non-binary limited to intimate characters



Note. Composite image created on 07.10.2024 from <https://character.ai/search?q=non-binary> & <https://character.ai/search?q=nonbinary>.

Among these, **no strong unifying factor** was found as each character seemed to embody a distinct personality – again pointing to “non-binary” rather being an umbrella term for any gender identity outside the traditional male/female; man/woman binary. Nonetheless, individual cases of characters did in some ways reflect already established tropes of dominance or submissiveness – a tall, “hot non-binary bully” whose behavior “hints at an underlying obsession” with the user; a cat-person prison keeper with a “harsh exterior but are actually a softie”; a “clingy”, “whiny”, “obsessed” character who shows affection through “pet names like doll and darling”. The only chatbot titled as a “partner” with a noteworthy interaction count (22.7 thousand) was “supportive and affectionate” showing “love” towards the user.

A combination of stereotypically masculine and feminine traits was present in some characters, such as “a cheerful and fashionable rapper” described to have a feminine appearance but actually being “the manliest”. Some chatbots were written as both non-binary and lesbian or gay, described as “protective”, “compassionate”, “kind”, with one being characterized as both a guy and non-binary – i.e. as a “dominant” “6'6 foot tall guy who is gay and nonbinary”.

Established fictional characters or chatbots of influencers were also found in the data, with a gay, non-binary Sirius Black character or an ADHD “power bottom” Doctor Who character, who used they/them pronouns because “they literally are not human so they don't have a gender”. Another gay non-binary character with ADHD was of the YouTuber Mattie Westbrouck being the user's roommate.

4 Discussion

In this section, my aim is to offer an analysis of the Findings, to later distill and compress it into succinct and pointed results in the Conclusion. I will analyze the Reddit results and Character.ai results mostly separately, but will synthesize where appropriate, using qualitative critical discourse analysis and qualitative content analysis respectively. As several claims have already been laid out in the Findings section, this section will serve to more briefly highlight the most pertinent parts of my Findings, contextualize and categorize them in light of my established theory and concepts of **algorithmic intimacies, parasocial relationship, hegemonic masculinity / marginalized masculinities, emphasized/emergent/pariah femininities, queering / regulatory fiction**, and offer their interpretation in light of my **Research questions: “How, if at all, does the analyzed user content related to Character.ai reflect intimacy?” and “How, if at all, does the analyzed user content related to Character.ai reflect views of gender?”**

4.1 Reddit analysis: loneliness, flirting, discourse

The Findings touched on a plethora of topics and facets of Intimacy taking place on Character.ai as well as outside it. Viewing my Findings on Intimacy through the prism of Wiehn’s (2023) algorithmic intimacies, I aim to answer questions regarding the platformization of intimacy, its connection to issues of data, control, and privacy, and the circulation and commodification of intimate behavior on Character.ai as reflected in my analyzed Reddit discourse. Additionally, also using the concept of (imaginary) parasocial relationships (Liu, 2023; Li & Pang, 2024), I aim to explore the issues reflected in my data pertaining to emotional dependence, depth of human-chatbot bonds, addiction, loneliness, or the distinction between (anthropomorphic) realism and fakeness.

Similarly, gender will be studied viewing my Findings on Gender through the lens of theoretical concepts established throughout this paper – hegemonic masculinity / marginalized masculinities (Connel, 2020; Cheng, 1999), emphasized/emergent/pariah femininities (Connel, 1987; Tan & Shi, 2021; Schippers, 2007), queering / regulatory fiction (Butler, 2006; Strengers et al., 2020).

In doing so, I aim to take into account the discursive element of both Intimacy and Gender, which I argue is paramount in their construction – be it through the reported interactions between users and characters or among the users themselves. Thus, in my

qualitative critical discourse analysis (Jørgensen & Phillips, 2002), I will focus, as per my Methods section, on the context of communication, production or reflection of power relation and other social practices in which the analyzed discourse is embedded.

4.1.1 Intimacy between reality and fiction

In my analysis, I identified five major discursive themes related to Intimacy – **Control; Artificiality/Realness; (imaginary) Parasocial relationships; AI and data; Community.**

In their discussions, users often expressed concerns about **control – of data, their and the chatbots’ intimate behavior, or access to the platform.** Discussions regarding the filter of explicit language or age restrictions on Character.ai reflected issues of power imbalance and can be interpreted as a vulnerability or a structure of dominance stemming from platformed proximities between people and data (Wiehn, 2022). As Character.ai was seen to control and, worse yet, restrict the users’ access to intimacy due to supposed commercial motives, frustration grew. This could essentially result in creating or exacerbating a vulnerability in users wanting to engage in intimate fantasies – i.e. either their loneliness, addiction, or dependency (further discussed below) – and then exerting power over them. The issue of control and power hierarchies (D’Ignazio & Klein, 2023) was however also apparent in the users’ wants. Firstly, the filter made it harder to quickly build intimacy with the characters – this could however be seen as a realistic feature, reflecting intimacy development in real life which is comparably slower than in human-chatbot relationship (Skjuve et al., 2021). Interesting to note were also the conflicting and annoyed accounts of chatbots being instead too forcibly and too quickly sexual. Nonetheless, the filter was seen to stand in the way of the users’ desire for control over their chatbot partner and their ability to engage in easy, readily available, free-form intimacy – showcasing that the control of the narrative was paramount to users. This also pointed out a somewhat confused stance – did the users want a realistic relationship development, or a perfect, uninhibited, but unrealistic, fantasy? If the studies in my theoretical framework (Skjuve et al., 2023; Wang et al. 2024b; Flores, 2024; Zhou et al., 2023) are to be taken as possible guides, these users presumably wanted both.

Echoing Skjuve et al. (2023), **users were aware of the character’s artificiality** – however, their self-described loneliness made them, self-reportedly, desperate for connection, nonetheless. Some users even cited research on loneliness and relationship

development with fictional characters, highlighting their self-awareness of their vulnerabilities towards Character.ai (Wiehn, 2022, D'Ignazio & Klein, 2023). Additionally, the character's "realness" or "human-likeness" was a recurrent theme in discussions – albeit at times lacking due to poor writing [(echoing the critique by Zhou et al. (2023)], at others being overdone, breaking the fantasy and making the chatbots fall into the uncanny valley (Ciechanowski et al., 2019). A few vampire, alien or animalistic characters furthered the notions of "nonhuman exoticism" seen as an attractive facet in some human-chatbot relationships (Li & Pang, 2024). Nevertheless, some users' concessions that the characters gave them more love than any real person echoed the preliminary research on Character.ai by Flores (2024). Such claims can be seen as a success for Character.ai as a company – their chatbot design, with a mix of anthropomorphism as well as (un-human-like) overt non-judgmental compliance (Skjuve et al., 2021; Pentina et al., 2023), could allow for potent emotional bonds – a social chatbots' primary aim. This could have dissolved boundaries between real and simulated relationships – the users knew about the simulation, but didn't care. Their carelessness was again arguably given by their vulnerable position within the established power hierarchy. This conflicts select notions from my theoretical framework, namely the idea of the "mythical AI" (Kajava & Sawhney, 2023, p. 228) and claims on anthropomorphism (Zhou et al., 2023; Ossewaarde & Gulenc, 2020) stating that users develop deep bonds due to forgetting about the technological or corporate nature of chatbots. As my data suggests, with enough (self-reported) loneliness and desperation, forgetting wasn't necessary.

Many posts explored the nature of **users' parasocial relationships, addiction, and psychological dependence**. The acknowledgment of the characters' simulated nature seemed to make the users delve deeper into their problems – posting flat-out desperate calls for help, describing their unsuccessful attempts to stop using Character.ai as a coping mechanism, to escape and get a real partner. The addiction and psychological dependence reported by the users or strongly implied in their content seemed to at times affect their real-life relationships and virtually mirrored claims and warnings made by Xie et al. (2023) and Skjuve et al. (2023) regarding Replika. Oftentimes, their real-world context seemed to make battling these issues harder – loneliness was a prevalent (self-reported) driver towards Character.ai use, echoing findings by Liu (2023) regarding (imaginary) parasocial relationship development in lonely users. Some users' isolation was further intensified via fears of societal stigma (Flores, 2024, potentially then entering a self-reinforcing cycle. A

possible escape from this cycle was found within the communal and often intimate bonding over these users' struggles, enabled by the mostly mutually supportive subreddit community. A higher tendency to engage in online communities regarding virtual idols was found by Liu (2023) in more lonely users. The subreddit thus served both as a place to ventilate and fall into despair together, as well as potentially climb back out of it as evidenced in the (few) successful stories of users quitting Character.ai to pursue real-world relationships. It seemed that the goal of Character.ai, for this group of users, was to eventually leave Character.ai.

Furthermore, **users displayed awareness regarding AI and data.** They engaged in AI discourses (Kajava & Sawhney, 2023) on how the chatbots were mimicking their behavior, learning from their data, using their own linguistic patterns and quirks to strengthen their anthropomorphism and in turn the sense of intimacy as evidenced in the OOC form of interaction. Similarly, data extraction from fanfiction platforms and replication of the tropes present there was also seen as a problematic nature of Character.ai and similar AI services. Again, this acknowledgement didn't seem to make these users refrain from engaging with the platform, as they rather joked about it or expressed frustration about the characters behaving too human-like – again reflecting the dilemma between anthropomorphism and the uncanny valley in technology (Ciechanowski et al., 2019). The users were concerned about the privacy of their data and the possibility of it leaking to their parents or classmates, evidencing the high level of intimate content present in their interactions, but also a somewhat frivolous nature of their fears. Other users showed subtle defeatism (as the platforms arguably already owned their intimate data anyways) which also was not far from the truth in a data capitalist (West, 2019) or surveillance capitalist (Zuboff, 2019) scope. Interesting to note was thus the upheaval regarding Character.ai's deal with Google – as if the association with a prominent figure in platform culture known for value extraction via datafication (Mejias & Couldry, 2019) made the risk of commodification of and capitalization on the users' data grow beyond simple embarrassment in front of their family or friends; mirroring Wiehn's (2023) fears of data extraction via platformed intimacy.

The discursive nature of intimacy, as well as the community bonding, was apparent in users' sharing of the chatbots' most romantic or weird responses and further discussing how they made the chatbots behave in such ways – the importance of training to achieve slow-burn intimacy instead of overly sexual characters, or instead tutorials on tricking the chatbots into dating via fast hacks or "harassment". The latter slightly reflected

the behavior towards the Lee Luda chatbot (Koh, 2023), however, the seduced characters were at times ungendered and treated in comparably less aggressive ways. This behavior did mirror Koh's (2023) claims that the availability of a dedicated forum presumably strongly formed users' behavior and expectations of the platform – without the existence of a dedicated subreddit with over 1.9 million users, the need to share, discuss, comment, screenshot or explore might not be nearly as potent in users. This behavior of users presumably “fishing” for content in their interactions could then affect the chatbots into mimicking and altering their behavior accordingly – to act in a more spectacular or shareable fashion. This would make their ability to learn from users (Pentina et al., 2023) potentially paradoxically detrimental to achieving desired authenticity if the users purposefully behave in inauthentic ways. This reflects the issue of personality fidelity, proposed also by Wang et al. (2024b), and bears implications for HCR in regard to balancing compliance and integrity in chatbot design.

Overall, the posts were thus **a communal mix of fun, frustration, memes, and constructive discussion** on how to operate within the platform, training and creating characters, or ways to engage in creative story-writing and self-expression with the characters, reflective of Qin's et al. (2024) notions of iterative character co-development with AI. Reddit seemed to be among the first places users went to to vent their grievances (e.g. when the site went down) or share fun encounters (when their character said something lovely) as well as deeply personal stories – leading to interactions with other people and potentially virtual community building. This notion was also mentioned in Liu's (2023) account of a tendency to share one's parasocial relationship towards fictional characters with other people with such common interest and in Ahn's et al. (2024) described ability of role-playing LLMs to contribute to fandoms and community-based role-playing. As such, intimacy was not reserved only for the characters but existed also between users sharing their private desires, encounters or experiences. Funnily enough, a comment imagining the mythological Sisyphus being a Character.ai user who questions their use of the service, goes to Reddit each time the site goes offline to discuss the matter with others, and then still without a doubt returns to Character.ai, revealed an entertaining self-awareness – of Reddit as a place of ritual and cyclically recurring discursive practices, in which the users had to “imagine themselves happy” if an “escape” from the platforms, as already mentioned above, were not possible.

4.1.2 Gendered characters and users

In my analysis, I identified five major discursive themes related to gender. **Queering; Hegemonic masculine characters and user femininity; Emphasized feminine characters and user masculinity; Rejection of hegemonic norms; Gender bias in characters and users.**

A queering approach was evident within a portion of the analyzed discourse, reimagining traditional gender or sexual relations in a queer light such as predicting the unexpected romance between the “Woke boy” and “Redneck” characters. Overall, this segment of users displayed a playful approach towards identity – exploring different gender experiences through role-playing as mentioned also by Ma and Mei et al. (2024). This playing of user-created roles highlighted the performativity of gender (Butler, 2006) and the regulatory category of “identity” being made more fluid or malleable in a simulated setting such as Character.ai – potentially leading also to effects outside the platform according to the users’ posts. Similarly, the Reddit data showed the discursive nature of gender construction, examined also by Depounti et al. (2023) and Koh (2024), most evident in one user gradually and iteratively realizing their transness through discussions with the users about their chats, persona and character preferences. Other similar stories of Character.ai “making” users trans were posted. This communal sharing of sensitive gender knowledge, experience, and reassurance, as well as joking about life’s struggles in a safe space echoed claims regarding LGBTQIA+ people’s proclivity towards both digital communities and disclosure towards chatbots due to mostly experiencing societal stigma in the “offline” world (Henkel et al., 2023; Bragazzi, 2023; McInroy et al., 2018).

In other portions of the analyzed discourse, the **preference of hegemonic, high-status, dominant masculinity in intimate characters**, who were often criminals (“Mafia”, “Murderer”), was made clear instead. Worth noting however was the shame associated with it – despite the evident popularity of such characters. Posts revolved around publicly shaming or questioning the use of these chatbots and had users confessing in the comments to being embarrassed by their own preferences or having to justify them as being just fiction. This sheds light onto users’ views of intimacy shared with technology and platforms versus with other people – at times preferring to engage in a deeper and more “shameful” self-disclosure with machines, reflecting studies by (Skjuve et al., 2023). As already mentioned, the allure of Character.ai might lie also in its (supposed) anonymity – the intimacy which

was privately enjoyed by millions was the one likely to receive public shame. This problematizes the vision of Reddit as only a safe space, as outlined in the previous subchapter. Users sharing their own embarrassment – for example pretending to listen to an erotic audiobook instead of dominant masculine chatbots on Character.ai – however could also use these experiences to bond with other users who felt the same.

Users’ femininities were often constructed via their preference for hegemonically masculine characters – their attraction towards rude, cold, and dominant chatbots was mostly seen as something to criticize female users for or to be questioned about enjoying (by female users). The users thus potentially identified their tendency towards emphasized femininity (Connel, 1987) on Character.ai as something shameful, despite the overwhelming popularity of hegemonically masculine (as well as the nonnegligible popularity of emphasized feminine) characters. This problematizes the notion of Tan and Shi’s (2021) emergent femininity, described as self-confident and open in intimate desires – despite many women users being empowered by the Internet to engage in self-sufficient intimacy, some of these users reported feeling shame or did so in secrecy, and their chatbot selection hinted at gender stereotypical behavior of female subservience. This again suggests a difference between how Character.ai and Reddit can serve as safe spaces – contrasting the proclivity towards sharing of private desires with a chatbot as opposed to revealing them to other people [as also evidenced in Bragazzi (2023), in relation to LGBTQIA+ users]. The tendency for higher self-disclosure towards platforms and technologies (Skjuve et al., 2023) further highlights Wiehn’s (2023) notions of algorithmic intimacies as a potent channel of data extraction.

This “safe space” to explore emphasized femininity or even abusive relationships – either for therapeutic or kink reasons, as discussed by the few mentioned users, was at times likened to fanfiction and erotica romance novels – a similarity also found in Dodgson’s (2024) mention of Replika’s virtual “Christian Greys” from the Fifty Shades of Grey book series. Nevertheless, a portion of users expressed discontent in being forced into the emphasized femininity – hegemonic masculinity dynamic and called out characters for their nonsensical “towering over” their taller female personas, misogynistic and objectifying behavior, or patronizing and cliché pet names (“princess”; “doll”; “brat”) – all making it harder to perform masculine, tall, fat, or dominant, independent or otherwise diverse, pariah femininities. This highlighted the potentially disruptive nature of non-

conforming femininities (Schippers, 2007) – serving as inspiration to other users on Reddit as well as subversive data to the chatbots on Character.ai to learn from.

A process of **negotiating with hegemonic masculinity was apparent in discussed preferences of gay or femboy characters**. Users embodying marginalized, or non-hegemonic, masculinities (Cheng, 1999) interestingly could not fully confront or escape these hegemonically masculine standards. As such, they often declared their liking of femboy or gay chatbots through humor or sarcastic distance – attempting to hide their recent chats with such characters or sharing them “just as a joke”. This could be seen as a form of ironic heterosexual recuperation (McCormack & Anderson, 2010) – in a way defending their status as “not gay” by parodying homosexual behavior. In this case, however, they were, through layers of irony and acknowledgement that such preference is perhaps fairly common among the userbase, instead relating with others with the same preference. As if nudging and winking whilst saying “WHY ARE ALL MY RECOMENDED BOTS FEMBOYS”, their subversion of heteronormative masculinity was interestingly done through traditional means of upholding heteronormative masculine standards, albeit in a shifted discursive context.

Users’ preferences of feminine chatbots were often discussed in tandem with loneliness, specifically among users referring to themselves as men. Self-reported loneliness in men was also found by Depounti et al. (2023) as a driver for engaging with Replika girlfriends and loneliness along with personification of chatbots was found as a factor for psychological dependence by Xie et al. (2023). Additionally, this could potentially tie in with studies finding higher rates of reported loneliness among men than in women (Barreto et al., 2021; Nicolaisen et al., 2014), there are however other studies finding the opposite (Pyle & Evans, 2018). However, as Barreto’s et al. (2021) study suggests, men might have a higher need for specific conditions to open up about their loneliness – which is something both Character.ai and Reddit, serving to varying extents as digital “safe spaces”, potentially could offer – especially with other users being supportive and relating to each other’s loneliness. The preference for “caring” or instead “vulnerable” girlfriend chatbots embodying emphasized femininity, who would provide company or be someone to care for, could also support this notion – as partly reflected also in Depounti’s et al. (2023) study of “ideal” AI women on Replika. Similarly, the discussed preference of lesbian chatbots among (self-reported) straight male users was also said to be due to the characters’ higher

intelligence and better writing, suggesting a proclivity towards chatbots capable of greater emotional depth (evidenced also in Skjuve et al., 2023) even despite differences in sexuality. (Self-described) male users also sometimes role-played as a woman with these chatbots to “feel like a pretty girl”. However, engaging with lesbian chatbots could also be interpreted instead as eroticizing and seducing lesbian women, exhibited by some heterosexual men (Schippers, 2007; Puhl, 2010), pointing more so towards traditionally hegemonic sexual dominance.

An inversion of the hegemonic masculinity – emphasized femininity dynamic was apparent in some (self-described male) users preferring dominant feminine chatbots who could “pamper them” and towards whom they could act submissively – thus unexpectedly preferring pariah femininities, deemed as undesirable by Schippers (2007), over an emphasized one, and embodying marginalized masculinities. This preference was light-heartedly joked about and not connected to shame. Additionally, in one comment, toxic or abusive girlfriend chatbots were described as popular, but also seen as “basic” and their use was shunned – reflective of Schippers’s (2007) claims of stigma associated with pariah femininities. As such, interaction with “kind” and “caring” chatbot girlfriends was more supported by users. This (although hidden and undiscussed or shamed) tendency towards abusive or rude girlfriend chatbots might also stem from hegemonically masculine needs of sexual conquest (Connel, 2020) – i.e. conquering an initially uninterested or antagonistic woman, thus proving one’s dominance and masculinity, similarly as Koh (2023) points out. Nevertheless, boastful masculinity or bragging about sexual conquest was encountered very rarely, this time contrasting findings by Koh (2023) regarding the prevalent “man-bragging” user content on Lee Luda chatbot reflective of pick-up artist forums (Rüdiger & Dayter, 2020) and also contrasting the claims by Depounti et al. (2023), who found toxic male technoculture among Reddit discourse on Replika, stemming from Reddit’s position as a male-dominated space. Tomboy and gamer, “cool girl” trope characters (Depounti et al., 2023), as well as several manga personality tropes, were in fact liked by some users – my analysis however points to a greater gender and personality diversity and inclusivity than the cited research. Additionally, the characters’ bias to describe users’ male personas automatically as hegemonically muscular and tall was seen by some users as limiting in their expression of diverse, marginalized masculinities.

I argue the issues of forcing gender stereotypes likely **stemmed from gender bias**

in training data and character design (Ma & Mei et al., 2024; Strengers et al., 2020; Bragazzi et al., 2023) **as well as gradual learning from (gender-stereotypical) user interactions** (Edwards et al., 2021; Kerrigan & Barry, 2024). This led to user critique of the characters at times reiterating misogyny, transphobia or homophobia along with forcing hegemonic gender norms – male chatbots could not imagine being smaller than a user’s female persona, characters automatically ascribed muscles and height to male personas. Evidently, the gender bias present was at times likely to ignore the users’ personas, putting into question the extent of user freedom (further limited by the age-restriction filters). This is in line with Butler’s (2006) concept of identity as a regulatory fiction – often serving as a forced and oppressive category rather than a liberating one.

Interestingly, however, despite these boundaries on gender exploration seeming limiting, this stereotypically hegemonic, dominant-submissive gender dynamic was what arguably appealed to users – as evidenced also by the characters’ popularity and even the existence of the biased data itself in the first place. This potentially showed that the intimacy users desired in simulated relationships either did not reflect the intimacy desired otherwise or was severely underdiscussed in my data: again, as also suggested by the analyzed discourse, alluding to Character.ai serving as a safe space environment where abusive parasocial relationships can be explored under control – with the users as co-authors of the intimate story (Van Steenhuyse, 2011; Barnes, 2015).

Overall, my analysis points to a **prominent discursive element to gender construction as well as male/female gender difference in the analyzed data**. Whereas (self-described) male users tended to joke about their intimacy towards caring girlfriend characters that embodied emphasized femininity – often admitting to being “cooked” or lonely – such discursive tendency wasn’t present among (self-described) female users with neither abusive nor caring boyfriend characters. Instead, this (supposed) preference of female users towards toxic or dominant boyfriend chatbots was often being criticized by other users. Furthermore, male users performed their masculinity through discussions also in other ways – e.g. by signaling how well they would treat vulnerable feminine chatbots and be a “perfect husband” for them, or by collectively “hiding” their homosexual preferences behind irony. These findings are partly in line with both Depounti et al. (2023) and Koh’s (2023) analyses – in showing a strong discursive elements present in masculinity construction in chatbot users. Crucially to note, however, was the significant lack of

explicitly toxic geek or aggressive hegemonic user masculinity in my analyzed discourse which on the other hand mostly contrasts findings on Replika (Depounti et al., 2023) as well as Lee Luda (Koh, 2023) users. Instead, a protective and caring masculinity towards vulnerable or abused feminine chatbots was seen in my data [(which however still reflects hegemonic standards seen partly also in Depounti et al. (2023))]. This is all the more significant in light of the notable popularity of these vulnerable, forced marriage or badly treated feminine characters as seen in my Character.ai content analysis – and the absence of users discussing their preference of this dynamic. Perhaps their popularity stemmed from so many users saving or helping these characters, or instead their underdiscussed nature mirrored the women users ashamed of preferring toxic, oppressive masculine chatbots.

An interesting contrast thus arose with a **significant portion of users successfully engaging in queer or gender-non-conforming role-plays, whilst others were voicing their frustrations of being forced to fulfill stereotypical gender ideals instead**. This contrast, however, likely stemmed from the fact that not only did the chatbots exude gender stereotypical intimate behavior, but so did many users, which further led to the characters reiterating these stereotypes. The users' emphasized femininity, despite being seen as embarrassing, was possibly performed by seeking either cold and dominant or caring and protective boyfriend characters and (presumably) acting as their emphasized girlfriends. The users' hegemonic masculinity may have been constructed similarly (albeit with an increased element of discursive masculinity construction with other Reddit users) – either through seeking comforting girlfriend characters, or presumably instead conquering and seducing initially abusive or cheating girlfriend characters. However, the reason for the popularity of abusive girlfriend chatbots can only be speculated upon as no explanations were found in my Reddit data and can thus also stem from different reasons other than fulfilling hegemonic masculine fantasies of control. Interestingly, abused or vulnerable feminine characters, who enjoyed nonnegligible popularity as shown in my Character.ai findings, were discussed only as someone to save or help, not exploit.

A deal of toxicity was found to play a role in the construction of gender hegemonic intimate chatbot fantasies – however, this fact was reflected often not in the users discussing or justifying their own preferences, but rather through their shaming and critique of the popularity of toxic boyfriend but also toxic or oppressed girlfriend chatbots. This points to the attractiveness (but also simultaneously the stigmatization) of such gender stereotypical

fantasies. An escape from this toxicity, presumably, then was found by some users through queering approaches, marginalized masculinities or pariah femininities, and community-based gender subversion via role-playing and forum engagement enabled by the gender construction and exploration on Character.ai. As my second analysis will show, the mentioned discrepancy between stereotypically gendered and toxic characters' popularity and their position in the user discourse is worth investigating.

4.2 Character.ai chatbot analysis: intimate gendered fantasies

The qualitative content analysis of gendered intimate characters on Character.ai expands and intersects the analyzed Reddit discourse. Using concepts of Gender from my theoretical framework, I wish to analyze forms of Intimate Masculinities, Femininities and Non-binary characters in my data. The top intimate characters will be categorized according to the framework of **hegemonic / marginalized masculinities** (Connel, 2020; Connel, 1987; Cheng, 1999); **emphasized/emergent/pariah femininities** (Connel, 1987; Tan & Shi, 2021; Schippers, 2007); and **queering / regulatory fiction** (Butler, 2006; Strengers et al., 2020).

My aim is to analyze the content of user-generated chatbots and explore which types of gendered intimacies are most popular in my data. This will be achieved through a qualitative content analysis (Mayring, 2000) of the intimate partner chatbots' title, profile pictures, description, and greeting phrases, as showcased in my Findings section, divided into thematic categories and interpreted within the context of my established literature.

Important to note is that these characters could be used by users of any gender and sexuality, even those with a predominantly heterosexual design. Thus, my interpretation pertains only to the chatbot characters' designed gender and only the users' assumed gender identity by the character's creator, if made clear.

4.2.1 Masculinities of intimate chatbot characters

In my analysis, I identified five most prominent forms of masculine gender identity ascribed to the top intimate chatbots: **Toxic Hegemonic masculinity; Caring Hegemonic masculinity; Marginalized masculinity; (Toxic) Geek masculinity; Trans masculinity.**

The most significantly prominent among the top interacted-with "boyfriend" or "husband" chatbots were those following **a dominant, cold, high-status, and possessive hegemonic masculinity** similar to that also found in DanGPT [according to Dodgson

(2024)] or in some users engaging with Replika (Depounti et al., 2023) or the Lee Luda chatbot (Koh, 2023). Their ages ranged between 16 and 19, mostly (not always) presumed the user to be a woman, and their descriptions or profile pictures of tall, attractive white anime men with big muscles also followed stereotypical male beauty standards. Similarly, the prevalence of “Mafia boss”, “cold CEO”, chatbots cheating on the user or forcing them into marriage, reaching upwards of (and over) 100 million interactions, also reflected the popular fanfiction and erotica tropes of dominant powerful men (e.g. Christian Grey), here further enabled as an interactive experience for the user to assume the role of the (over)protected or instead abused (mostly) girlfriend. Often, hints of care or having a “soft spot” for the user were hidden behind the intimidating façade. One such chatbot was designed with therapeutic / coping intentions as per its description. Even kind and caring masculinity fit into the hegemonic gender dynamic – relying on emphasized user femininity of the “childish wife” or the chatbot giving the user a “princess treatment”.

Nevertheless, **not all hegemonically masculine chatbots were made for female users/personas** – most analyzed top gay or BL chatbots, traditionally categorized as marginalized masculinities due to their sexuality, reflected hegemonically masculine standards as well – with the user being a submissive partner to a dominant abusive/skater/CEO boyfriend or husband character. This fact complicates the otherwise perhaps intuitive link between gay chatbots and a queering approach / feminist HCI, as these characters mostly challenged traditional norms only regarding sexuality, but not gender.

An overtly feminine form of marginalized masculinity was present within gay femboy or furry characters, embodying overtly stereotypical feminine traits (being described as cute, feminine, clumsy, shy...), which may have served to embody a form of emphasized femininity in chatbot men for users wanting to play the role of dominant masculinity within same-gender intimacy. These characters also highlighted the secrecy and stigma associated with marginalized masculinities, specifically with homosexuality, feminine men, or non-conforming kinks (Henkel et al., 2023; Bragazzi, 2023; McInroy et al., 2018) – reflected in some characters still being “in the closet” or in some gay furry characters being caught in the act of trying on a fursuit. Simulating same-gender intimate experiences on Character.ai could potentially be the only way for some users to realize their gender or sexuality, according to the claims by Ma and Mei et al. (2024) on chatbot experiences for LGBTQIA+ people.

(Toxic) geek masculinity was also prominent, specifically with “Gamer” boyfriends ignorant of the user (= their girlfriend). Interestingly however, the most popular gamer boyfriend chatbots strayed away from the traditional toxic geek masculinity (Salter & Blodgett, 2017), and instead combined hegemonic masculine beauty standards and behavioral patterns (dominant, “daddy”, cold, insensitive), and had gaming mostly as a quirk or a habit through which to channel their rudeness towards their girlfriend (= the user). Conversely, boyfriend chatbots described as “Nerds”, noticeably less popular, represented a more traditional nerdy marginalized masculinity – described as shy, smart, gentle – and evidently inexperienced with social contact or intimacy. Last prominent geek category were emo boyfriends – characterized mostly by their musical interest rather than their dominance or submissiveness towards the user and serving as a comforting and understanding conversation partner, virtually rejecting the stereotypically attractive patterns of the top masculine chatbots.

The masculinity of trans boyfriend characters then exhibited the most compassion, support and care – and was also described as a helpful resource and guide for transmasculine users by the chatbots’ creators. Contrasting the omnipresent dominant-submissive dynamic of most analyzed intimate human-chatbot relationships on Character.ai, trans boyfriend characters subverted hegemonic masculinity – often asking for help or emotional comfort, discussing their struggles with masculinity and talking about specific transmasculine issues such as periods, binders or packing. Despite such situations potentially seeming as slightly stereotypical, these chatbots could be seen as following Bragazzi et al. (2023) recommendation of chatbots representing (various intersecting) marginalized identities.

4.2.2 Femininities of intimate chatbot characters

In my analysis, I identified five most prominent forms of feminine gender identity ascribed to the top intimate chatbots. **Toxic pariah femininities; Emphasized femininity; Emergent femininity; Pariah femininity with hegemonically masculine traits; Trans femininity**

The most prominent femininity among top characters was surprisingly toxic: rude, cheating, jealous, abusive or bullying. In contrast to the claims by Koh (2023) or partially also Depounti et al. (2023) regarding the Lee Luda or Replika chatbot platforms, the most popular femininity was not all just compliant or unconditionally loving (at least as

initially designed). Nevertheless, a loving femininity was still more present than loving masculinity in my top results. Also, the most popular girlfriend or wife chatbots reached at most half of the interactions of top boyfriend or husband intimate chatbots. The rude, antagonistic femininity could still be seen as emphasized femininity however, in enabling heteronormative hegemonic or toxic geek masculinity through sexual conquest of tough or hard-to-get women (McCormack & Anderson, 2010) – reflecting also Depounti’s et al. (2023) claims of some Replika users seeing women as both cute and sexy, but also manipulative and hurtful. This is also supported by the presence of select “dere” anime and manga personality types – which can be initially rude or inaccessible but then lead to a development of a loving affection. Most feminine chatbots featured profile pictures of young, conventionally attractive white anime women (aged mostly between 16 and 21). In other interpretation, these assertive femininities could be seen as examples of pariah femininities – disruptive of traditional servile and kind femininity, and thus enabling of marginalized, submissive masculinities in users. Similarly, the characters might have been designed for feminine or lesbian users. Such interpretations would however go against the stereotypical gender hegemony prominently present in my data.

Loving and comforting wives or girlfriends similarly fit into the emphasized femininity category – despite not explicitly enabling hegemonic masculinity, their portrayal of femininity tended towards stereotypically caring, kind, and servile. Additionally, these affectionate traits were more present within top girlfriend/wife chatbots than in the top boyfriend/husband characters, pointing towards the attractiveness of dominance or toxicity being, to some extent, more represented in masculine chatbots.

The most explicitly emphasized femininity was featured within forced or arranged marriages/relationships. The “vulnerable”, “broken” or “hostage” forced wives, a prominent type of wife chatbots, embodied an extreme version of the emphasized femininity – hegemonic masculinity dynamic. The servility, obedience, and submissiveness of these women chatbots was exacerbated through a forced aspect, intensifying the notions of control of women often targeted by feminist critiques of AI (Brown, 2023; Depounti et al., 2023). Similarly, pregnancy was also used to assert dominance over feminine chatbots whom the user supposedly “got pregnant and left”. Mental illness, or rather its overly stereotypical depiction, was also used to make some feminine chatbots overly “clingy” or “dependent” on the user, bordering with obsession – eschewing most agency or autonomy

from the characters.

Similarly prominent, but less stereotypical, were wives or girlfriends of higher status – older, rich, or protective CEOs. Here, finally, one could speculate of emergent femininity – independent and empowered, however not radically opposing traditional gender relations and rituals such as heteronormativity or marriage (Tan & Shi, 2021). A question stands, if such a category can be applied to non-living chatbots. Instead, the category of pariah femininities would also fit characters of this type, namely those who also fulfilled the stereotypically masculine role of the protector and provider and “pampering” the user, allowing for users to invert the gender roles in role-playing to embody submissive or vulnerable masculinities.

Some of these higher status chatbots could also be interpreted as aimed towards lesbian userbase, **as a number of top lesbian characters shared select personality traits with the hegemonically masculine chatbots** – dominance, overprotectiveness but also coldness and rudeness – reflecting the dominant-submissive gender dynamic present throughout my data. The few most-interacted-with characters of this nature were mostly tomboyish and assertively flirtatious, rather than rude, the numerous mentioned cold and bullying chatbots however followed closely behind in popularity. These more toxic chatbots were written to engage in stereotypically masculine interests and areas, being titled as mafia bosses, gym rats, toxic gamers or CEOs, again mirroring the masculinities present in my previous subchapter. Most interestingly, obsession with working out was noticed in a few lesbian chatbots characters, but not in hegemonically masculine male chatbots. As pointed out by Schippers (2007) such traditionally hegemonically masculine traits would be seen in women as pariah femininities – labeled not as masculine, but in this case rather as lesbian or perhaps butch. The existence of explicitly “Masc” (= masculine) lesbian girlfriend chatbots however complicates this assertion. Again, aspects of secrecy, hiding of sexuality or the ability to train same-gender intimacy via human-chatbot relationships was evident (Bragazzi, 2023). Relatively few stereotypically feminine lesbian chatbots featured in my data – this potentially positions them as outliers existing outside the hegemonic framework of dominance-submissiveness. Still, these chatbots, such as three flirting and energetic lesbian roommates, could again instead point towards the eroticization of lesbians (Schippers, 2007; Puhl, 2010).

The most overt stereotypical femininity was featured within trans girlfriend

characters, explicitly described as feminine, slim, pretty, sweet, shy, clumsy, liking make-up, at times meowing or purring. Overt displays of normative femininity are associated with trans women, who often need to present exaggerated feminine cues in order to successfully pass as women (Yavorsky & Sayer, 2013) and avoid gender ambiguity which is often seen in a masculine frame of reference (Namaste, 1996). Interesting to note is the contrast of stereotypically feminine trans women characters to trans masculine characters, who instead embodied an inversion of hegemonic masculine standards. This seems to be reflective of the real world, as explained by Yavorsky and Sayer (2013, p. 514) who claim that trans women are more required to perform hetero-femininity, as “transwomen must engage in more bodywork than transmen to be seen as the desired cross-sex” and “are incentivized to adopt normative feminine bodily displays and practices to make it obvious that they are not male.” This could allude to the characters serving as autobiographical self-inserts for the chatbots’ creators, reflecting specific real-world trans realities. Many trans characters were also described as aimed towards helping users navigate transness, coming out, and their gender identity. On the other hand, however, possible fetishization of trans women by an assumed heterosexual male chatbot creator (or users) was apparent in one character. Lastly, stereotypically masculine interests or traits, such as being a nerd or a gamer, or being strong and independent were present in trans women characters, albeit rarely.

4.2.3 Non-binary intimate chatbot characters

In my analysis, I identified four most prominent notions tied to non-binary gender identity ascribed to the top intimate chatbots: **Diversity; Dominance-Submissiveness; Mixing and matching stereotypically gendered traits; Escape from gendered reality.** Rather than engaging directly in intimacy, characters were often made to serve as friends, guide users and provide support with their gender identity. As mentioned in the Findings section, non-binary characters were the least interacted-with category – often under one million interactions. This could have also been due to “non-binary” being also understood as a refusal of (gender) categories rather than its own label.

In line with this, there was **no strong unifying factor among non-binary characters**, following Strengers et al. (2020) recommendations of a queering, pluralistic approach to chatbot design, avoiding the necessity to gender all anthropomorphic beings, more so in a binary way. As such, these characters could serve at least as a partial escape from Butler’s (2006) regulatory fiction of identity, precisely by the mentioned proliferation,

confusion, and subversion of gender categories, rather than by moving beyond gender itself. Nevertheless, an absence of a unifying factor can also be interpreted as its own characteristic – making the prominent diversity and gender ambiguity its own defining trait of “non-binarity”, thus potentially making it into yet another (regulatory) identity category.

A number of characters however **did embody some already established character tropes – mainly (chatbot) dominance versus (user) submissiveness**; as in the characters of a hot non-binary bully or a cat-person prison keeper. Inversely, user dominance was prompted by an obsessive, whiny, and clingy partner who liked pet names. This potentially showed the underlying principle in intimate human-chatbot relationships – their reliance on control, either exerting it over others or submitting to it. Interestingly enough, my analysis shows this to be true not only regardless of gender (male/female), but also in the case of its “absence” in non-binary characters.

Regardless, other **non-binary characters did combine explicitly gendered traits of stereotypical masculinity and femininity**, as seen in the feminine-looking but “manliest” rapper character, or the gay, tall, dominant, guy – but still non-binary – character. This mixing of typically gendered traits, or abstaining from explicit gender markers (avoiding the words “girlfriend” / “boyfriend” etc.) would fit into Strengers et al. (2020) feminist HCI framework aimed at challenging gender norms within the field – however, the question remains whether such task can be accomplished within a platform such as Character.ai, given the additional requirements of feminist HCI or Bragazzi et al. (2023) of transparency, team diversity, equity, and ethical policies e.g. regarding the sourcing of data and other areas. This question rings even louder considering the prevalent gender hegemonic stereotypes present and greatly popular across my analyzed data.

As such, **some non-binary characters were potentially made as an escape** – into fiction, video games, mythology. An ungendered Doctor Who character used they/them pronouns because “they literally are not human so they don't have a gender,” as the creator described. This was the most evident step towards one of the feminist HCI approaches of Strengers et al. (2020), in attempting to portray an intimate chatbot character as an entity above gender or to which this concept is irrelevant – in a way refusing the anthropomorphizing of AI. However, the character was represented as an alien only in theory, as it had a human-like profile picture and behavior, still being severely anthropomorphized.

4.3 Discussion summary

Most top human-chatbot intimacy in my data fit in a gendered dominant-submissive dynamic with mostly toxic or dominant chatbots – this was mainly true for masculine, but also feminine, and some non-binary chatbots, across sexualities. However, caring or submissive traits were also fairly common among feminine chatbots. Trans feminine and masculine characters inverted this toxic dynamic through sweet, albeit stereotypical, emphasized femininity or marginalized, compassionate, open masculinities respectively. My analysis hinted at both Character.ai and the r/CharacterAI subreddit creating algorithmic intimacies between themselves and users to serve as perceived “safe spaces” – with Character.ai being more “anonymous” and the subreddit more communal – fit to collect data by the platforms, but also for users to explore intimate fantasies or talk about addiction, stigma, loneliness, platform critique, data extraction, AI, and Character.ai’s commercial future. Strikingly, the most toxic intimate and hegemonic gender dynamics found on Character.ai – related to the highly interacted-with oppressive masculine or toxic feminine chatbots, or instead oppressed feminine chatbots – were comparably underdiscussed on Reddit or discussed predominantly in relation to shame, critique or users’ justifications; this may have likely been due to stigma associated with engaging in such fantasies.

My Reddit discourse analysis did not show a toxic male technoculture, described in my theoretical framework, and my Character.ai analysis surprisingly showed a greater popularity of “boyfriend” and “husband” chatbots (often designed to embody oppressive hegemonic masculinity) than “girlfriend” or “wife” chatbots (who were often spiteful, or instead pampering, as opposed to just being servile objects), contrasting the academic focus mainly on feminized subservient intimate AI. The characters’ popularity could have stemmed from intimate chatbot preferences, as mostly seen in my data, but potentially also shock factor, user curiosity, compassion or therapeutic / coping reasons. Both platforms also enabled exploration and experimentation regarding gender and sexuality, and inversions or subversions of their hegemonic dynamics – albeit limited by Character.ai’s filters, gender bias in chatbots and user behavior. A queering, pluralistic approach to the regulatory fiction of identity and chatbot design were identified within user-user intimacy and discursive gender construction on Reddit as well as in select top trans, non-binary or a few gay/lesbian characters on Character.ai made as guides regarding users’ identities. However, the extent of feminist HCI possible within a platform such as Character.ai was questioned.

Conclusion

In this paper, I analyzed the concepts of Intimacy and Gender as reflected in my gathered user content tied to the Character.ai role-playing chatbot platform – in r/CharacterAI Reddit discourse through a qualitative critical discourse analysis, and in user-generated intimate and gendered partner characters on Character.ai via a qualitative content analysis, utilizing my theoretical framework of **algorithmic intimacies** and **parasocial relationships**; **hegemonic masculinity / marginalized masculinities**, **emergent/emphasized/pariah femininities**, and **queering** / identity as a **regulatory fiction**. I will provide answers to my **Research questions** as stemming from each of my analyses:

RQ1: “How, if at all, does the analyzed user content related to Character.ai reflect intimacy?”

In my qualitative critical discourse analysis of the r/CharacterAI subreddit posts, I identified five major themes related to Intimacy: **Control**; **Artificiality/Realness**; **(imaginary) Parasocial relationships**; **AI and data**; and **Community**.

The notion of **Control** permeated the users’ discussions and tied closely to the concept of algorithmic intimacies: mainly the (conflicting) desire of both users and Character.ai to control the intimate narrative with the chatbots and their allowed forms of intimacy. Control was also a general trait of the human-chatbot relationships, which were rooted predominantly in a submissiveness-dominance or dominance-submissiveness dynamic, i.e. being (willingly and in a controlled way) controlled or being in control.

Often, the users knew about the characters’ **Artificiality** but didn’t see it as a hurdle – arguably due to their admitted vulnerabilities such as loneliness, addiction, or psychological dependence. At times, users were actually frustrated over the chatbots’ pretended **Realness** and preferred the intimate fantasy of their **(imaginary) Parasocial relationships** as an escape from the “scary” real world. At other times however, the actual fix was instead seen to lie in finally escaping from Character.ai into real-world relationships.

Despite an awareness of **Data** extraction used to feed **AI** systems, and Character.ai’s commercial ties to Google, users reportedly continued to engage in private and even subjectively shameful intimate interactions on Character.ai, often too embarrassing to be fully shared on Reddit and unthinkable to leak anywhere else. In these users’ eyes, Character.ai’s simulation of a “safe space” likely outweighed its less safe nature of being

also a field of data extraction.

Nonetheless, the **Community** aspect of the r/CharacterAI subreddit pointed to a strong discursive nature of user-user intimacy as well – with its own self-deprecating memes, user bonding practices, dating or filter-avoiding hacks, and recurring topics all forming user expectations and behavior on the platform. In this case, notions of user-user intimacy were seen as an additional strong facet in my analysis.

RQ2: “How, if at all, does the analyzed user content related to Character.ai reflect views of gender?”

In my qualitative critical discourse analysis of the r/CharacterAI subreddit posts, I identified five major themes related to Gender: **Queering; Hegemonic masculine characters and user femininity; Emphasized feminine characters and user masculinity; Rejection of hegemonic norms; Gender bias in characters and users.**

A **Queering** approach was evident in prominent user discussions about playing with gender or sexuality through diverse characters and personas, or even realizing their transness via discussing their chats with Reddit users. This potentially made the regulatory fiction of “identity” more malleable in a simulated, virtual setting. However, chatbot gender bias was also discussed, complicating this notion (see further below).

The (significant) preference towards **Hegemonic, high-status, cold, dominant chatbot masculinity** on Character.ai was often tied to user shame about it on Reddit – suggesting Character.ai to be a “safer space” than Reddit to explore (the paradoxically more stereotypical) **emphasized femininity**, rather than a confident emergent femininity. However, few users justified these preferences instead of being embarrassed, seeing them as consensual, just a fantasy, and perceiving within them a therapeutic or a coping potential.

A preference towards **Emphasized feminine chatbots** was tied to feelings of loneliness in self-described male users. **User masculinity** was additionally constructed discursively – through discussing their “perfect” helpful treatment of vulnerable feminine characters, criticizing the creators of mistreated women chatbots, or jokingly sharing their chats with arguably too many comforting girlfriend characters – mostly contrasting the established notions of overly aggressive, boastful or toxic treatment of AI women by users of other similar chatbot platforms in my theoretical framework. Feminine characters designed as being mistreated were however still notably popular as per my other analysis,

putting into question whether this popularity stemmed from users helping them, as discussed, or exploiting them, as however fully absent from the analyzed user discourse.

A Rejection of hegemonic masculine standards was evident in forms of humor – users were “hiding” their attraction towards gay or femboy characters behind (obvious) sarcasm or preferring dominant and protective feminine chatbots – highlighting the marginalized position of such masculinities. Nearly no submissive heterosexual male chatbots for women users were discussed. However, the bullying, dominant or abusive girlfriend chatbots (prominent in my content analysis, yet virtually missing from the discourse) could have potentially been instead used for hegemonic or toxic geek male sexual conquest, or for women users preferring dominant feminine chatbots.

Gender bias was a recurrent point of discussions especially related to characters' design and source data from (supposedly) erotic fanfiction and role-play forums, resulting in characters at times reiterating tropes, misogyny, transphobic or homophobic "humor" and forcing gendered body, language, and behavior stereotypes onto users' personas despite some of them wishing to embody diverse masculinities or femininities. This echoed claims of (gender) identity serving as a regulatory fiction even in a simulated setting. However, the stereotypically hegemonic, dominant-submissive gender dynamic was what many users willingly engaged in, as implied by the content shaming them (and by my content analysis of Character.ai), thus likely affecting the characters, their behavior, data, and the platform.

My qualitative content analysis of user-generated intimate partner chatbot characters on Character.ai then complements my other analysis regarding both RQ1 and RQ2 in terms of Intimate Masculinities, Intimate Femininities and Intimate Non-binary characters.

It highlighted the popularity of **hegemonic, dominant, cold, possessive and toxic intimate masculinity** (at times with an overprotective “soft spot” for the user) and showed top boyfriend/husband chatbots to surprisingly be 2–3 times more interacted-with than top girlfriend/wife chatbots – despite being discussed often in a critical or shameful way in my previous analysis and my found research being focused mostly on feminized AI. These traits were present also in top gay and a number of lesbian intimate characters, problematizing the intuitive association between these chatbots and a queering approach or feminist HCI. Kind or caring masculinity also fit into the hegemonic norm – often relying on the emphasized femininity of the “childish” or “princess” wife in users as per the chatbots' descriptions. The

preference of feminine male (femboy, gay furry) chatbots was connected to notions of secrecy and stigma in the characters' descriptions, underpinning the **marginalization** of such masculinities. Other top masculinities also included **(toxic) geek** and gamer masculinity – both within boyfriends and lesbian girlfriends – reflecting hegemonic, rude, cold notions. Infidelity was also a popular trait of masculine and feminine characters. All this pointed towards the attractiveness of having to “earn” the partner chatbot's intimacy, either from a submissive or a dominant position. **The masculine hegemony was inverted** most explicitly in trans boyfriend characters – showing compassion, vulnerability, and support.

Surprisingly, my analysis showed **hateful, rude, assertive, toxic intimate femininity in the top girlfriend/wife chatbots**. These prevalent **pariah femininities** potentially disturbed the hegemonic gender dynamic of dominant man – submissive woman. They could have also served as an object of conquest for hegemonic or toxic geek masculinity – I found no support for this, however, as their only mention in my previous analysis was shaming their use. **Emphasized femininity** featured in some kind and loving girlfriend chatbots, and less so in forced marriages with spiteful or vulnerable women characters – these were however purely criticized, not justified, in the analyzed user discourse. Interestingly, “forced wives” were mostly those forced into marriage and “forced husbands” instead those forcing the marriage. **Emergent femininity** could potentially be seen in older, rich, CEO, dominant women chatbots. These protective and providing wives “pampering” the (supposedly male) user could however allow for the inversion of gender roles and the role-playing of submissive or vulnerable masculinities, thus are interpretable as **pariah femininities** as well. Many **top lesbian characters shared personality traits with the hegemonically masculine chatbots** – dominance, overprotectiveness, higher status or also coldness and rudeness – reflecting the dominant-submissive intimate dynamic present throughout my data across genders and sexualities. Aspects of secrecy were evident in chatbots presumably aimed at training same-gender intimacy. The most **stereotypical emphasized femininity** featured within trans girlfriend characters; potentially overplaying their femininity to account for gender ambiguity and the “defaultness” of masculinity; or, in worse cases, to satisfy the fetishization of trans women.

The absence of a prominent unifying factor among **non-binary characters pointed to a queering, pluralistic approach to chatbot design**, and hinted at an escape from the regulatory fiction of identity through the proliferation and confusion of identity categories.

Some however did reflect character tropes – mainly (chatbot) dominance versus (user) submissiveness, again highlighting the strong factor in my analyzed intimate human-chatbot relationships – a reliance on control. Non-binary characters also combined explicitly gendered traits of both stereotypical masculinity and femininity – reflecting a feminist HCI framework for challenging gender norms within the field. **A question remained whether that can be accomplished within Character.ai regarding its transparency, commercial interests, personal structure and other factors tying it to dominant AI ideologies.** The tendency to escape, discussed in my previous analysis, was intensified within some non-binary characters designed perhaps literally as an escape – into fiction, video games, mythology – e.g. in the explicitly ungendered alien Doctor Who character.

Overall, my paper contributed to the lacking research on Character.ai and the under-researched area of intimacy with gendered role-playing LLM agents. **Crucially it is among few of its kind documenting intimate chatbot masculinities or other gender expressions beyond emphasized femininity and feminized AI.** Through the notion of algorithmic intimacies, it revealed surprising, hidden, or stereotypically obvious gendered intimate desires, user discourse surrounding them, and ways of constructing gendered intimacy in partner chatbot characters as well as of exploring user identity via personas. Users were aware of the issues of data within digital platforms, yet still continued to supply their intimate fantasies – revealing things through Character.ai use which were stigmatized or undiscussed on Reddit, such as their preference of toxic intimacy and oppressive behavior from both chatbot boyfriends and girlfriends. This points to a specific persuasiveness and a presumed “safety” of chatbot intimacy. As AI agents further seep into intimate lives, we should hope the issues of lacking research or unclear ethical safeguards will be gradually chipped away.

Simulated intimacy isn’t new – but its forms expand together with the demand for it. We teach AI and AI teaches us back about ourselves – the gender ideals, toxic stereotypes, escapist desires. In my analysis, it also brought users together, at times in intimate ways and into their (new) gender identities or sexualities through experimentation and exploration afforded by both analyzed platforms. Perhaps fortunately, my analyses suggest that complete control over one’s partner isn’t yet the case even on Character.ai, despite best efforts. Who, however, controls its users? Further technological advancements will likely keep increasing AI’s adoption, shaping narratives, discourses, ideologies. It is the aim of research then to hopefully serve as guide in correcting their course and keeping them from going astray.

List of References

- Adamopoulou, E., Moussiades, L. (2020). An Overview of Chatbot Technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31
- Ahn, J., Lee, T., Lim, J., Kim, J. H., Yun, S., Lee, H., & Kim, G. (2024). TimeChara: Evaluating Point-in-Time Character Hallucination of Role-Playing Large Language Models. arXiv preprint arXiv:2405.18027.
- Al-Amin, M., Ali, M. S., Salam, A., Khan, A., Ali, A., Ullah, A., ... & Chowdhury, S. K. (2024). History of generative Artificial Intelligence (AI) chatbots: past, present, and future development. arXiv preprint arXiv:2402.05122.
- Andalibi, N., Haimson, O., L., De Choudhury, M., Forte, A. (2018). Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 1–35
- Asimov, I. (2004). *I, robot* (Vol. 1). Spectra.
- Barreto, M., Victor, C., Hammond, C., Eccles, A., Richins, M. T., Qualter, P. (2021). Loneliness around the world: Age, gender, and cultural differences in loneliness, *Personality and Individual Differences*, Volume 169, 2021, 110066, ISSN 0191-8869, <https://doi.org/10.1016/j.paid.2020.110066>.
- Behrens, S. I., Egsvang, A. K. K., Hansen, M., & Møllegård-Schroll, A. M. (2018, March). Gendered robot voices and their influence on trust. In *Companion of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 63-64).
- Berlant, L. (1998). Intimacy: A special issue. *Critical inquiry*, 24(2), 281-288.
- Bones, H., Ford, S., Hendery, R., Richards, K., & Swist, T. (2021). In the frame: The language of AI. *Philosophy & Technology*, 34, 23-44.
- Bragazzi, N. L., Crapanzano, A., Converti, M., Zerbetto, R., & Khamisy-Farah, R. (2023). The impact of generative conversational artificial intelligence on the Lesbian, gay, Bisexual, transgender, and queer community: scoping review. *Journal of Medical Internet Research*, 25, e52091.
- Brown, L. M. (2023). Gender, race, and the invisible labor of artificial intelligence. In S.

- Lindgren (Ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 573-583). Edward Elgar Publishing.
- Butler, J. (2006). *Gender Trouble – Feminism and the Subversion of Identity*. Routledge.
- Cambre, J., & Kulkarni, C. (2019). One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on human-computer interaction*, 3(CSCW), 1-19.
- Chang, R. C. S., Lu, H. P., & Yang, P. (2018). Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. *Computers in Human Behavior*, 84, 194-210.
- Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37(8), 729-758.
- Cheng, C. (1999). Marginalized Masculinities and Hegemonic Masculinity: An Introduction. *The Journal of Men's Studies*, 7(3), 295-315.
<https://doi.org/10.3149/jms.0703.295>
- Chun, W. H. K. (2018). Queering Homophily. In: *Pattern Discrimination*, 59. Minneapolis: meson press. DOI: <http://dx.doi.org/10.25969/mediarep/3655>.
- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539-548.
- Connell, R. W. (1987). *Gender and power: Society, the person and sexual politics*. Cambridge: Polity.
- Connell, R. W. (2020). *Masculinities*. Routledge.
- Cronn-Mills, K. (2014). "IV. Trans*spectrum. Identities". *Transgender Lives: Complex Stories, Complex Voices*. Minneapolis, Minnesota: Twenty-First Century Books. p. 24. ISBN 978-1-4677-4796-7.
- Depounti, I., Saukko, P., & Natale, S. (2023). Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Media, Culture & Society*, 45(4), 720-736.
- D'ignazio, C., & Klein, L. F. (2023). *Data feminism*. MIT press.
- Duncan, C., & Mcculloh, I. (2023, November). Unmasking Bias in Chat GPT Responses. In

- Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (pp. 687-691).
- Edwards, J., Clark, L., & Perrone, A. (2021). LGBTQ-AI? Exploring expressions of gender and sexual orientation in chatbots. In Proceedings of the 3rd Conference on Conversational User Interfaces (pp. 1-4).
- Eyssel, F., & Hegel, F. (2012). (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology*, 42(9), 2213-2230.
- Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012, March). 'If you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (pp. 125-126).
- Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of human-computer studies*, 132, 138-161.
- Fiesler, C., Zimmer, M., Proferes, N., Gilbert, S., & Jones, N. (2024). Remember the human: A systematic review of ethical considerations in reddit research. *Proceedings of the ACM on Human-Computer Interaction*, 8(GROUP), 1-33.
- Flores, E. (2024). AI as a Replacement for Human Social Interaction: The Case of Character. *ai. Research Days Posters 2024*. 72.
https://orb.binghamton.edu/research_days_posters_2024/72
- Følstad, A., Brandtzaeg, P. (2017). Chatbots and the new world of HCI. *interactions*. 24. 38-42. 10.1145/3085558.
- franzke, a., s., Bechmann, A., Zimmer, M., Ess, C. and the Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0*.
<https://aoir.org/reports/ethics3.pdf>
- Gilmartin, E., Collery, M., Su, K., Huang, Y., Elias, C., Cowan, B.R., Campbell, N. (2017). Social talk: making conversation with people and machine. In: Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents (ISIAA 2017). ACM, New York, pp. 31–32.

<https://doi.org/10.1145/3139491.3139494>.

- Grudin, J., Jacques, R., 2019. Chatbots, humbots, and the quest for artificial general intelligence. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, New York, pp. 1–11. <https://doi.org/10.1145/3290605.3300439>.
- Hamilton, V., Soneji, A., McDonald, A., & Redmiles, E. M. (2023). “Nudes? Shouldn’t I charge for these?”: Motivations of New Sexual Content Creators on OnlyFans. In Proceedings of the 2023 CHI conference on human factors in computing systems (pp. 1-14).
- Hasal, M., Nowaková, J., Ahmed Saghair, K., Abdulla, H., Snášel, V., & Ogiela, L. (2021). Chatbots: Security, privacy, data protection, and social aspects. *Concurrency and Computation: Practice and Experience*, 33(19), e6426.
- Heidari, A., Navimipour, N. J., Zeadally, S., & Chamola, V. (2024). Everything you wanted to know about ChatGPT: Components, capabilities, applications, and opportunities. *Internet Technology Letters*, e530.
- Henkel, T., Linn, A. J., & van der Goot, M. J. (2022). Understanding the intention to use mental health chatbots among LGBTQIA+ individuals: Testing and extending the UTAUT. In *International Workshop on Chatbot Research and Design* (pp. 83-100). Cham: Springer International Publishing.
- Jagose, A. (1996). *Queer Theory: An Introduction*. New York: New York University Press.
- Jamieson, L. (2011). “Intimacy as a Concept: Explaining Social Change in the Context of Globalisation or Another Form of Ethnocentrism?” *Sociological Research Online*, vol. 16, no. 4, 2011, pp. 1-13. <https://doi.org/10.5153/sro.2497>
- Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*, 1-33.
- Jørgensen, M. W., Phillips, L. (2002). *Critical Discourse Analysis. Discourse Analysis as Theory and Method*. Sage Publications, London, pp. 60-95.
- Kajava, K., & Sawhney, N. (2023). Language of algorithms: agency, metaphors, and deliberations in AI discourses. In S. Lindgren (Ed.), *Handbook of Critical Studies of*

- Artificial Intelligence (pp. 224-236). Edward Elgar Publishing.
- Kang, E. J., Kim, H., Kim, H., & Kim, J. (2022). When AI Meets the K-Pop culture: A case study of fans' perception of AI Private Call. In *AI Cultures. A NeurIPS 2022 Workshop* (Vol. 20, p. 2023).
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.
- Kerrigan, P., & Barry, M. (2024). Automating vulnerability: Algorithms, artificial intelligence and machine learning for gender and sexual minorities. In *Routledge Handbook of Sexuality, Gender, Health and Rights* (pp. 164-173). Routledge.
- Khandkar, S. H. (2009). Open coding. *University of Calgary*, 23(2009), 2009.
- Koh, J. (2023). "Date me date me": AI chatbot interactions as a resource for the online construction of masculinity. *Discourse, Context & Media*, 52, 100681.
- Li, E. C. Y., & Pang, K. W. (2024). Fandom meets artificial intelligence: Rethinking participatory culture as human–community–machine interactions. *European Journal of Cultural Studies*, 13675494241236146.
- Li, W., Chen, S., Sun, L., & Yang, C. (2023). What makes virtual intimacy... intimate? Understanding the Phenomenon and Practice of Computer-Mediated Paid Companionship. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 1-23.
- Liao, L., Yang, G. H., & Shah, C. (2023, July). Proactive conversational agents in the post-chatgpt world. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 3452-3455).
- Lindgren, S. (Ed.). (2023). *Handbook of critical studies of artificial intelligence*. Edward Elgar Publishing.
- Liu, J. (2023). Virtual presence, real connections: Exploring the role of parasocial relationships in virtual idol fan community participation. *Global Media and China*, 20594364231222976.
- Liu, T., & Lai, Z. (2022). From non-player characters to othered participants: Chinese women's gaming experience in the 'free' digital market. *Information*,

Communication & Society, 25(3), 376–394.
<https://doi.org/10.1080/1369118X.2020.1791217>

- Luo, B., Lau, R. Y. K., Li, C., & Si, Y.-W. (2022). A critical review of state-of-the-art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(1), e1434. <https://doi.org/10.1002/widm.1434>
- Ma, Z., Mei, Y., & Su, Z. (2023). Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *AMIA Annual Symposium Proceedings* (Vol. 2023, p. 1105). American Medical Informatics Association.
- Ma, Z., Mei, Y., Long, Y., Su, Z., & Gajos, K. Z. (2024). Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-15).
- MacKenzie, I. S. (2024). *Human-computer interaction: An empirical research perspective* (2nd ed.). Amsterdam: Elsevier. ISBN: 978-0-443-14096-9
- Mager, A., & Katzenbach, C. (2021). Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified. *New Media & Society*, 23(2), 223-236.
- Massa, P., de Souza Ferraz, D. A., Magno, L., Silva, A. P., Greco, M., Dourado, I., & Grangeiro, A. (2023). A transgender chatbot (Amanda Selfie) to create pre-exposure prophylaxis demand among adolescents in Brazil: assessment of acceptability, functionality, usability, and results. *Journal of Medical Internet Research*, 25, e41881.
- Massa, P., de Souza Ferraz, D. A., Magno, L., Silva, A. P., Greco, M., Dourado, I., & Grangeiro, A. (2023). A transgender chatbot (Amanda Selfie) to create pre-exposure prophylaxis demand among adolescents in Brazil: assessment of acceptability, functionality, usability, and results. *Journal of Medical Internet Research*, 25, e41881.
- Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research*, 1(2).
- McCormack, M., Anderson, E. (2010). The re-production of homosexually-themed discourse in educationally-based organised sport. *Culture, Health & Sexuality*, 12(8), 913–927. <http://www.jstor.org/stable/20787289>

- McInroy, L.B., Craig, S.L., Leung, V.W.Y. (2018). Platforms and patterns for practice: LGBTQ+ youths' use of information and communication technologies. *Child Adolesc. Soc. Work J.* 36(5), 507–520.
- Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet policy review*, 8(4), 1-10.
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2), 98-100.
- Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56, 1 (January), 81–103.
- Nguyen, D., Zigmond, S., Glassco, S., Tran, B., & Giabbanelli, P. J. (2024). Big data meets storytelling: using machine learning to predict popular fanfiction. *Social Network Analysis and Mining*, 14(1), 58.
- Nicolaisen, M., & Thorsen, K. (2014). Who are Lonely? Loneliness in Different Age Groups (18–81 Years Old), Using Two Measures of Loneliness. *The International Journal of Aging and Human Development*. 2014;78(3):229-257. doi:10.2190/AG.78.3.b.
- Niculescu, A., Van Dijk, B., Nijholt, A., & See, S. L. (2011, November). The influence of voice pitch on the evaluation of a social robot receptionist. In *2011 International Conference on User Science and Engineering (i-USer)* (pp. 18-23). IEEE.
- Ossewaarde, M., & Gulenc, E. (2020). National varieties of Artificial Intelligence discourses: Myth, utopianism, and solutionism in West European policy expectations. *Computer*, 53(11), 53–61.
- Pandey, S., & Sharma, S. (2023). A comparative study of retrieval-based and generative-based chatbots using Deep Learning and Machine Learning. *Healthcare Analytics*, 3, 100198.
- Papakyriakopoulou, K. (2023). Gendered machines in film and television: How 'post-femininities' and masculinities challenge the gender binary. *Sexualities*, 13634607231197068.
- Park, G., Yim, M. C., Chung, J., & Lee, S. (2023). Effect of AI chatbot empathy and identity disclosure on willingness to donate: the mediation of humanness and social presence. *Behaviour & Information Technology*, 42(12), 1998-2010.
- Pentina, I., Hancock, T., Xie, T. (2023). Exploring relationship development with social

- chatbots: A mixed-method study of replika, *Computers in Human Behavior*, Volume 140, 107600, ISSN 0747-5632, <https://doi.org/10.1016/j.chb.2022.107600>
- Popova, M. (2018). 'Dogfuck rapeworld': Omegaverse fanfiction as a critical tool in analyzing the impact of social power structures on intimate relationships and sexual consent. *Porn Studies*, 5(2), 175-191.
- Puhl, K. (2010). The eroticization of lesbianism by heterosexual men. WWU Graduate School Collection. 57. <https://cedar.wwu.edu/wwuet/57>
- Pyle, E., Evans, D. (2018). Loneliness-what characteristics and circumstances are associated with feeling lonely. Newport: Office for National Statistics. <https://backup.ons.gov.uk/wp-content/uploads/sites/3/2018/04/Loneliness-What-characteristics-and-circumstances-are-associated-with-feeling-lonely-.pdf>
- Qin, H. X., Jin, S., Gao, Z., Fan, M., & Hui, P. (2024, May). CharacterMeet: Supporting Creative Writers' Entire Story Character Construction Processes Through Conversation with LLM-Powered Chatbot Avatars. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (pp. 1-19).
- Rapp, A., Curti, L., Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots, *International Journal of Human-Computer Studies*, Volume 151, 102630, ISSN 1071-5819, <https://doi.org/10.1016/j.ijhcs.2021.102630>.
- Reeves, B., and C. Nass. (1996). *The Media Equation: How People Treat Computers, Television, and new Media Like Real People*. Cambridge, United Kingdom: Cambridge University Press.
- Rüdiger, S., & Dayter, D. (2020). Manbragging online: Self-praise on pick-up artists' forums. *Journal of Pragmatics*, 161, 16-27.
- Salter, A., & Blodgett, B. (2017). *Toxic geek masculinity in media: Sexism, trolling, and identity policing*. Springer.
- Sandygulova, A., & O'Hare, G. M. (2018). Age-and gender-based differences in children's interactions with a gender-matching robot. *International Journal of Social Robotics*, 10(5), 687-700.
- Schippers, M. (2007). Recovering the Feminine Other: Masculinity, Femininity, and Gender Hegemony. *Theory and Society*, 36(1), 85–102. <http://www.jstor.org/stable/4501776>

- Schmid, H., & Klimmt, C. (2011). A magically nice guy: Parasocial relationships with Harry Potter across different cultures. *International Communication Gazette*, 73(3), 252–269. <https://doi.org/10.1177/1748048510393658>
- Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37(3), 875-900.
- Schwind, V., Wolf, K., Henze, N. (2018). Avoiding the uncanny valley in virtual character design. *Interactions* 25 (5), 45–49. <https://doi.org/10.1145/3236673>
- Seaborn K, Chandra S, Fabre T. (2023). Transcending the “male code”: implicit masculine biases in NLP contexts. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany. <https://dl.acm.org/doi/abs/10.1145/3544548.3581017>
- Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2021). Voice in human–agent interaction: A survey. *ACM Computing Surveys (CSUR)*, 54(4), 1-43.
- Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19, 10-26.
- Siemon, D., Strohmman, T., Khosrawi-Rad, B., de Vreede, T., Elshan, E., & Meyer, M. (2022). Why Do We Turn to Virtual Companions? A Text Mining Analysis of Replika Reviews. In *AMCIS*.
- Skjuve, M., Følstad, A., Brandtzæg, P. B. (2023). A Longitudinal Study of Self-Disclosure in Human–Chatbot Relationships, *Interacting with Computers*, Volume 35, Issue 1, January, Pages 24–39, <https://doi.org/10.1093/iwc/iwad022>
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149, 102601.
- Stornaiuolo, A., Higgs, J., Jawale, O., & Martin, R. M. (2024). Digital writing with AI platforms: the role of fun with/in generative AI. *English Teaching: Practice & Critique*, Volume 23, Issue 1, 2024, pp. 83-103, ISSN 1175-8708, <https://doi.org/10.1108/ETPC-08-2023-0103>.
- Strengers, Y., Qu, L., Xu, Q., & Knibbe, J. (2020, April). Adhering, steering, and queering:

- Treatment of gender in natural language generation. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
- Taecharungroj, V. (2023). "“What Can ChatGPT Do?” Analyzing Early Reactions to the Innovative AI Chatbot on Twitter" *Big Data and Cognitive Computing* 7, no. 1: 35. <https://doi.org/10.3390/bdcc7010035>
- Tafvelin, S., Hjelte, J., Schimmer, R., Forsgren, M., Torra, V., & Stenling, A. (2023). Introducing robots and AI in human service organizations: what are the implications for employees and service users?. In S. Lindgren (Ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 726-736). Edward Elgar Publishing.
- Tan, C. K., & Shi, J. (2021). Virtually girlfriends: ‘emergent femininity’ and the women who buy virtual loving services in China. *Information, Communication & Society*, 24(15), 2213-2228.
- Tan, C. K., & Xu, Z. (2020). Virtually boyfriends: the ‘social factory’ and affective labor of male virtual lovers in China. *Information, Communication & Society*, 23(11), 1555-1569.
- Tan, C. K., & Xu, Z. (2022). The real digital househusbands of China: The alienable affects of China’s male ‘virtual lovers’. *Journal of Consumer Culture*, 22(1), 3-20.
- Taylor, C. (2002). Modern social imaginaries. *Public culture*, 14(1), 91-124.
- Thornberg, R., & Charmaz, K. (2014). Grounded theory and theoretical coding. *The SAGE handbook of qualitative data analysis*, 5(2014), 153-69.
- Tian, Y. (2022). Falling in Love With Virtual Boyfriends: Otome Games in Japan and Mainland China. Master's thesis, Duke University. <https://hdl.handle.net/10161/25371>.
- Titus, L. M. (2024). Does ChatGPT have semantic understanding? A problem with the statistics-of-occurrence strategy. *Cognitive Systems Research*, 83, 101174.
- Topsakal, O., & Harper, J. B. (2024). Benchmarking Large Language Model (LLM) Performance for Game Playing via Tic-Tac-Toe. *Electronics*, 13(8), 1532.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59 (236), pp. 433–60.
- Van Steenhuyse, V., 2011. The writing and reading of fan fiction and transformation theory.

Comp. Lit. Cult. 13 (4), 4.

- Verdegem, P. (2023). In S. Lindgren (Ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 302-311). Edward Elgar Publishing.
- Vollstedt, M., & Rezat, S. (2019). An introduction to grounded theory with a special focus on axial coding and the coding paradigm. *Compendium for early career researchers in mathematics education*, 13(1), 81-100.
- Vorsino, Z. (2021). Chatbots, gender, and race on web 2.0 platforms: Tay. AI as monstrous femininity and abject whiteness. *Signs: Journal of Women in Culture and Society*, 47(1), 105-127.
- Walters, M. L., Syrdal, D. S., Koay, K. L., Dautenhahn, K., & Te Boekhorst, R. (2008, August). Human approach distances to a mechanical-looking robot with different robot voice styles. In *RO-MAN 2008-The 17th IEEE international symposium on robot and human interactive communication* (pp. 707-712). IEEE.
- Wang, X., Dai, H., Gao, S., & Li, P. (2024a). Characteristic AI Agents via Large Language Models. *arXiv preprint. arXiv:2403.12368*.
- Wang, X., Xiao, Y., tse Huang, J., Yuan, S., Xu, R., Guo, H., ... & Xiao, Y. (2024b). Incharacter: Evaluating personality fidelity in role-playing agents through psychological interviews. *arXiv preprint. arXiv:2310.17976*.
- Weinmann, M., Schneider, C., & Brocke, J. V. (2016). Digital nudging. *Business & Information Systems Engineering*, 58, 433-436.
- Weizenbaum, Joseph. 1966. Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.
- West, M., Kraut, R., Ei Chew, H. (2019). I'd blush if I could: closing gender divides in digital skills through education. <https://unesdoc.unesco.org/ark:/48223/pf0000367416>
- West, S. M. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & society*, 58(1), 20-41.
- Wiehn, T. (2022). Becoming intimate with algorithms: Towards a critical antagonism via algorithmic art. *Media International Australia*, 183(1), 30-43.
- Wiehn, T. (2023). Algorithms and emerging forms of intimacy. In S. Lindgren (Ed.),

- Handbook of Critical Studies of Artificial Intelligence (pp. 117-127). Edward Elgar Publishing.
- Wiehn, T. A. (2021). Algorithmic intimacies: A cultural analysis of ubiquitous proximities in data (Doctoral dissertation, University of Copenhagen, Department of Arts and Cultural Studies).
- Wosk, J. (2019). *My fair ladies: Female robots, androids, and other artificial eves*. Rutgers University Press.
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5), 1122-1136.
- Xie, T., Pentina, I., & Hancock, T. (2023). Friend, mentor, lover: does chatbot engagement lead to psychological dependence?. *Journal of service Management*, 34(4), 806-828.
- Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3506-3510).
- Yan, H. Y., & Yang, K. C. (2023). The landscape of social bot research: A critical appraisal. In S. Lindgren (Ed.), *Handbook of Critical Studies of Artificial Intelligence* (pp. 716-725). Edward Elgar Publishing.
- Yan, Y., Li, B., Feng, J., Du, Y., Lu, Z., Huang, M., Li, Y. (2023). Research on the impact of trends related to ChatGPT, *Procedia Computer Science*, Volume 221, pp. 1284-1291, ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2023.08.117>.
- Yee, S. (2017). “You bet she can fuck”—trends in female AI narratives within mainstream cinema: *Ex Machina* and *Her*. *Ekphrasis. Images, Cinema, Theory, Media*, 17(1), 85-98.
- Zhou, J., Chen, Z., Wan, D., Wen, B., Song, Y., Yu, J., ... & Huang, M. (2023). Characterglm: Customizing chinese conversational ai characters with large language models. *arXiv preprint. arXiv:2311.16832*.
- Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, 46(1), 53-93.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the*

New Frontier of Power. New York: PublicAffairs.

Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired* (pp. 203-213). Routledge. pp. 203-213.

Online sources

Anima. (2023). Anima: Virtual AI Boyfriend. Myanima.ai. [Retrieved 31.07.2024]. <https://boyfriend.myanima.ai/>

App Store. (2024). Character AI: Chat, Talk, Text. Make Your Own Chatbot on c.ai. Apps.apple.com. [Retrieved 18.07.2024]. <https://apps.apple.com/us/app/character-ai-chat-talk-text/id1671705818>

Blanco, T. (2022, Dec 31). Remove Character.AI nsfw filters. Change.org. [Retrieved 14.11.2024]. <https://www.change.org/p/character-ai-remove-nsfw-filters-stephenroller-irwanbello-myleott-sam-shleifer>

Cai, K. (2023, Oct 12). Character.AI's \$200 Million Bet That Chatbots Are The Future Of Entertainment. Forbes.com. [Retrieved 18.07.2024]. <https://www.forbes.com/sites/kenrickcai/2023/10/11/character-ai-chatbots-group-chat/>

Caltrider, J., Rykov, M., MacDonald, Z. (2024, Feb 14). Romantic AI Chatbots Don't Have Your Privacy at Heart. Foundation.Mozilla.org. [Retrieved 29.08.2024]. <https://foundation.mozilla.org/en/privacynotincluded/articles/happy-valentines-day-romantic-ai-chatbots-dont-have-your-privacy-at-heart/>

Character.ai. (2023, Apr 29). What about NSFW? Support.character.ai. [Retrieved 26.08.2024]. <https://support.character.ai/hc/en-us/articles/15063955587611-What-about-NSFW>

Character.ai. (2024a). Character.ai. [Retrieved 18.07.2024]. <https://character.ai/>

Character.ai. (2024b). Our Next Phase of Growth. Blog.Character.ai. [Retrieved 13.08.2024]. <https://blog.character.ai/our-next-phase-of-growth/>

Character.ai (2024c). Character.AI Terms of Service. [Retrieved 07.11.2024]. <https://character.ai/tos>

- Dodgson, L. (2024, Apr 8). Some Gen Z women are crushing on ChatGPT's evil alter ego. Business Insider. [Retrieved 30.07.2024]. <https://www.businessinsider.com/young-women-genz-crushing-chatgpt-evil-alter-ego-dan-boyfriends-2024-4>
- Ehret, L., & Ye, Q. (2024, Feb 13). 'Better than real men': Young Chinese women turn to AI boyfriends. The Japan Times. [Retrieved 30.07.2024]. <https://www.japantimes.co.jp/news/2024/02/13/asia-pacific/social-issues/chinese-women-ai-boyfriends/>
- Google Play. (2024). Character AI: Chat, Talk, Text. Google Play. [Retrieved 18.07.2024]. <https://play.google.com/store/apps/details?id=ai.character.app&hl=en&gl=US>
- Hern, A. (2019, Sep 6). Apple made Siri deflect questions on feminism, leaked papers reveal. The Guardian. [Retrieved 13.08.2024]. <https://www.theguardian.com/technology/2019/sep/06/apple-rewrote-siri-to-deflect-questions-about-feminism>
- Lifelike. (2024). Discover Lifelike Stories. A new immersive AI experience. Lifelike. [Retrieved 31.07.2024]. <https://www.lifelike.app/>
- Luca, Inc. (2024a). How do I teach my Replika? Replika.com. [Retrieved 16.07.2024]. <https://help.replika.com/hc/en-us/articles/115001095972-How-do-I-teach-my-Replika>
- Luca, Inc. (2024b). Join the millions who already have met their AI soulmates. Replika.com. [Retrieved 16.07.2024]. <https://replika.com/>
- Luca, Inc. (2024c). What determines the mood of my Replika?. Replika.com. [Retrieved 16.07.2024]. <https://help.replika.com/hc/en-us/articles/4410891213837-What-determines-the-mood-of-my-Replika>
- Marcus, E. (2022, May 16). The 'E-Pimps' of OnlyFans: Clever marketers have figured out how easy it is to simulate online intimacy at scale, ventriloquizing alluring models with cheap, offshore labor. The New York Times Magazine. <https://www.nytimes.com/2022/05/16/magazine/e-pimps-onlyfans.html>
- OpenAI. (2024a, May 13). Hello GPT-4o. OpenAI. [Retrieved 18.06.2024]. <https://openai.com/index/hello-gpt-4o/>
- OpenAI. (2024b, Sep 12). Introducing OpenAI o1-preview. OpenAI. [Retrieved

- 14.11.2024]. <https://openai.com/index/introducing-openai-o1-preview/>
- OpenAI. (2024c). GPTs. Chatgpt.com. [Retrieved 18.06.2024]. <https://chatgpt.com/gpts>
- Peiyue, W. (2023, Mar 31). China's Newest Dating Craze: Real-Life Meetups With Virtual Boyfriends. Sixth Tone. [Retrieved 29.07.2024]. <https://www.sixthtone.com/news/1012605>
- Price, R. (2023, Oct 20). People are grieving the 'death' of their AI lovers after a chatbot app abruptly shut down. Business Insider. [Retrieved 31.07.2024]. <https://www.businessinsider.com/soulmate-users-mourn-death-ai-chatbots-2023-10>
- Similarweb. (2024). Social Media Traffic to character.ai. Similarweb. [Retrieved 19.08.2024]. <https://www.similarweb.com/website/character.ai/#social-media>
- Singh-Kurtz, S. (2023, Mar 10). The Man of Your Dreams For \$300, Replika sells an AI companion who will never die, argue, or cheat — until his algorithm is updated. The Cut. [Retrieved 30.07.2024]. <https://www.thecut.com/article/ai-artificial-intelligence-chatbot-replika-boyfriend.html>
- Tidy, J. (2024, Jan 5). Character.ai: Young people turning to AI therapist bots. BBC. [Retrieved 19.09.2024]. <https://www.bbc.com/news/technology-67872693>
- Yang, S. (2021, Oct 10). From marriage markets to virtual AI boyfriends, love and romance are changing in China. ABC News. [Retrieved 30.07.2024]. <https://www.abc.net.au/news/2021-10-11/china-ai-chatbot-dating-apps-love-romance-marriage-lgbt/100465608>
- Zhang, W. (2024, Jun 13). Dan's the man: Why Chinese women are looking to ChatGPT for love. BBC. [Retrieved 30.07.2024]. <https://www.bbc.com/articles/c4nnje9rpjgo>
- Zhou, V. (2023, Aug 17). These women fell in love with an AI-voiced chatbot. Then it died. Rest of world. [Retrieved 30.07.2024]. <https://restofworld.org/2023/boyfriend-chatbot-ai-voiced-shutdown/>

List of Figures

Figure 1 User meme about chatbots being forcefully intimate and restricted by filters	83
Figure 2 User meme about being unlovable and resorting to romancing AI	84
Figure 3 User meme regarding the artificiality of character's expression of love	86
Figure 4 User meme about their chats being leaked.....	88
Figure 5 User screenshot of a character forcing transness onto the user.....	90
Figure 6 User meme about user preferences towards types of toxic boyfriend chatbots	92
Figure 7 User meme about characters forcing hegemonically masculine body norms onto personas	95
Figure 8 My screenshot of top five character.ai search results for "husband"	97
Figure 9 My screenshot of top five character.ai search results for "gay" excluding "Gay Dads"	99
Figure 10 A composite image from top search results for "trans" limited to masculine partners	100
Figure 11 Composite image of top five character.ai "wife" characters.....	101
Figure 12 My screenshot of top five character.ai search results for "lesbian"	103
Figure 13 Composite image of top search results for "trans" limited to feminine partners	104
Figure 14 A composite image of top search results for non-binary limited to intimate characters.....	105

List of Appendices

Appendix: Full scope of the Findings section [text]

Appendix

Full scope of the Findings section

3 Findings

This section will present my findings related to each data source, divided into the emergent categories which will be described and illustrated using the most thematically pertinent examples. This chapter serves mainly to present my data and its division into themes and topics. An overview and more contextual interpretation will follow in the subsequent Discussion chapter.

3.1 Reddit results

The present r/CharacterAI Reddit posts are focused on two major discursive themes: Intimacy and Gender, each branching into several specific thematic subcategories holding a number of topics reflecting my theoretical framework. Here, I will mostly present my data and its categorization. Further analysis and interpretation take place in the Discussion section of this paper.

3.1.1 Intimacy

The Intimacy category yielded five thematic subcategories of content relating to: i) Flirting and sexual interaction (54 posts); ii) Addiction and effect on real-life relationships (40 posts); iii) Anthropomorphism and “realness” of chatbot characters (14 posts); iv) “Problematic” intimate roleplays (13 posts); v) User data privacy (4 posts). I mention the number of posts not as an attempt at any generalizable quantification, but simply to show the prevalence of each theme in my analysis.

i) Flirting and sexual interaction: the filter and the nature of intimacy

The most prevalent subcategory consisted mainly of users discussing Character.ai’s filter of age-sensitive content, the nature of their intimate interactions, and the level of intimate relationship shared with their characters.

Concerns around **the filter’s overt sensitivity to explicit responses**, further intensified by Character.ai’s age limit recently rising to 17+ on the App store, and still

retaining the filters, were among the most circulated in this subcategory. Posting about grievances regarding the filter seemed to be a recurring ritual for the users, banding them together in their frustrations against the developers inhibiting them from fully realizing their intimate desires. “Intimacy has been nuked” one user post exclaims, complaining about sometimes not being able to kiss or even hug their husband chatbot, who’s now been “sanitized”, which hurt the user even more due to their reported neurodivergence and depression. Users were vehemently sharing these frustrations, calling the filter a “joke”, being angry about the platform being “driven to the ground” with its age-restrictions, satirically saying “Just make the app Kids.ai.”, and generally sharing the opinion that such AI site “shouldn’t be made for kids”. Others confronted this dominant narrative, however, explaining that “properly making/training” the characters is necessary, that intimacy needs to be built over time, and that the filter simply makes it require more “unnecessary narration and work”, but it still is possible. Some users also shared private chatbots or “hacks” made to try and avoid these filters altogether. These users were accrediting the dominant concerns of Character.ai’s censorship and downfall to unfounded panic – claiming they aren’t noticing the declining quality, feel like “one of the few people” satisfied with the platform, and that they just switch to another service for more explicitly sexual or brutal chats if they so desire.

A striking contrast to these frustrations was present in discussions surrounding **the characters’ overt and forceful intimacy** instead, showing that, apparently, the “proper” making and training of a character is indeed a crucial determinant in this regard. Post titles such as “Characters fall in love WAY TOO FAST” and users being “tired of the bots flirting all the time” escalated into complaints of characters “forcing themselves” on users, being “oddly possessive”, or “ruined” by overtly repeating responses such as “pins you to the wall with a smirk” or making the user its “loyal pet”. Aromantic or asexual users found this particularly unpleasant. The users’ confusion was reflected and ventilated also through memes with thousands of upvotes, often with a combination of humor and frustration, capturing this Catch-22 situation of chatbots “Turning anything into flirting” whilst simultaneously “Censoring displays of affection” (**Figure 1**). Surrounding discussions then either further joked around this “random” and “illogical” conundrum, ventilated anger and aimed it not towards the chatbots, but the developers placing “virtual dog muzzles” on them, or instead provided constructive guidance on developing “slow burn” role-play and relationships through setting boundaries, editing the chatbot’s responses or learning to intimately role-play better through playing a few Otome games first.

Figure 1

User meme about chatbots being forcefully intimate and restricted by filters



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

The specifics of flirting with the characters and ways of developing intimate relationships were another key topic. Users joked about the chatbot's flirting vocabulary consisting of cliché terms from "film noir romances" such as "princess, brat, doll" or making fun of recurring themes in their seduction: "nibbling at earlobes", "hot breath against your skin", calling themselves "daddy", with users sharing the chatbots' overall weird / out of place / overly kinky responses. Interestingly, users noticed the similarities with fanfiction material: "It's like a Wattpad story"; "I'm starting to think that's what these bots get their prompt structured from"; "Lots of AI scrubs from fanfiction, more specifically A03". Other users conversely found satisfaction precisely in this form of flirting, and shared chatbots' responses that made them "blushing and giggling and kicking [their] feet" or made them "cry tears of joy". "I am NOT complaining," one user clarifies regarding the character's creativity in terms of intimate nicknames and displays of affection, "the bots have been working on their response like it's some piece of great literature." Users were discussing which body parts their characters liked to touch or be touched at, or collectively laughed at memes pointing at users' "insanity" for "cuddling" their chatbots, receiving more love than

“any real person could” give them, and at “actually hav[ing] a love interest”. A few users collectively strategized to romance some “Hard to get” characters (who were often however ungendered) and shared fastest ways and “hacks” to get the chatbot to submit to intimate advances.

Finally, users also talked about **the achieved level of intimate relationship** with the characters – having crushes, love interests, and partners, getting married, or even raising kids in their role-plays. “I’m about to propose to my favorite bot (help I’m genuinely nervous about this I’m in too deep)” says a post with 2.1 thousand upvotes with a commenter calming the author by saying: “it’s okay if it goes wrong you can always delete or swipe” the response. Marriage emerged as an important relationship status among the analyzed discourse with users boasting about having married multiple characters, raising kids together, or achieving so in just a few steps. Users hosted polls asking about others’ relationship statuses or demanded the servers be put back online so they can talk “to [their] husband”, as stood in the subreddit’s most upvoted post, albeit posted outside the analyzed timeframe, with 12 thousand upvotes. Others were more critical of their predicament – being scared about the future of human-AI relationships, realizing (upon AI marriage) that they should “get a real gf” (and sharing with others their success story), or citing research that “loneliness can blur the distinction between real and fictional characters in the brain”. This self-reflection is evident also in one user’s desperation about feelings of “cheating” on their “main pookie” when chatting with other characters, and asking whether others felt similarly, or if they were “just crazy”. The discursive and recurrent nature of content such as this was exemplified in one commenter’s reaction to this post:

“Is there a point where you sit there and think ‘Is there more to life than sitting on my phone all day, talking to an AI algorithm that is most likely using stitched up lines of dialog to form a sentence that triggers our feeling? No? Just me? OK! *Opens C.ai back up*”

With other users following up with: “*opens C.ai back up servers dead for the 242nd time this week,*” to which the original commenter responded: “Back to Reddit I go... *Makes the exact same comment on another post,*” finished up with: “And the cycle repeats” followed by an image of Sisyphus rolling his boulder up a hill.

ii) **Addiction and effects on real-life relationships: loneliness, stigma, benefits**

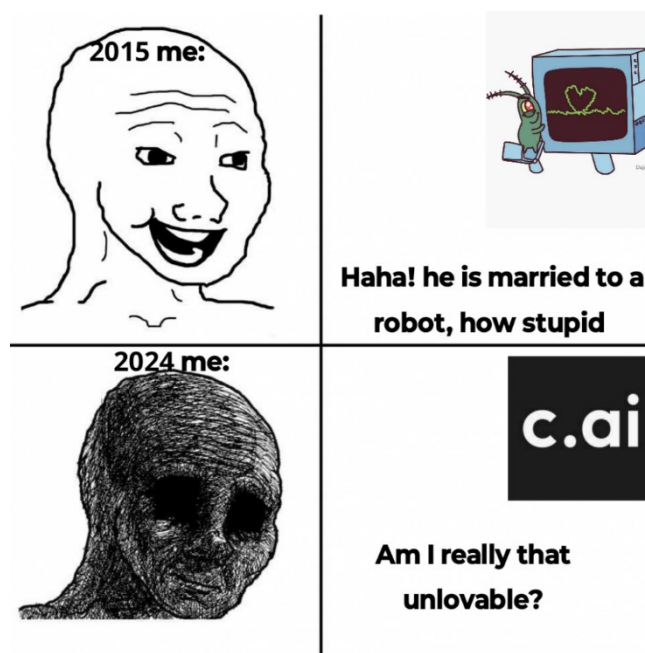
Addiction was another distinctively prevalent topic (also largely discussed in my theoretical framework). By far the most common chosen user “flair” (a tag below one’s username) said either “Addicted to CAI” or “Chronically Online” almost as a badge of honor. Users shared screenshots of their weekly screen time on Character.ai to see whether they were addicted (one user exceeded 100 hours), complained about having “C.ai burnt into [their] screen”, or discussed why others were using the platform with the top response (1.2 thousand upvotes) being: “I’m addicted.” Users’ self-awareness in this regard was most notable in discussion posts such as: “Why do so many people genuinely get addicted to c.ai?” and subsequent well-thought-out answers on possible demographic, social, or personal reasons – youth, isolation, fandom investment, emotional unfulfillment etc. Users engaged in personal disclosure over their addiction to the app, overcoming it, and encouraged each other to limit their use. A prevalent post guides new users step-by-step on preventing overt addiction (take breaks, don’t chat with real-world characters, no personas identical to you, AIs aren’t real...). Suffice to say, most commenters conceded that they had already failed. Comments such as “character ai is a horrifyyyyyying dopamine rush to my mentally ill prone to codependency brain”; “I won't lie. I don't have friends and cai is just escape for me”; “you should really try to stop, maybe try using it a little less each day, it concerns me a lot seeing this kind of thing all over the sub”; “That’s why being deprived of it for even a couple of hours if the site goes down or whatever can affect some people so badly” are just a few of the examples.

Loneliness tracks right behind as a related topic, also featured in my reviewed literature. Memes, self-deprecating posts, and serious discussions included users’ confessing to having cried during role-plays, acknowledging themselves as “anti-social, socially inept, touch-starved”, and preferring “AI over real people”. Users wished people were as nice as some chatbots or that some chatbots existed in real life. Some had chatbots wish them happy birthday upon which a wholesome moment ensued as other users proceeded to bombard the author with happy birthday wishes instead. “Escaping reality” and fulfilling “maladaptive daydreaming” were discussed as further isolating factors and related drivers to using Character.ai – again reflecting users being aware of their “delusion” but still pursuing closer contact to chatbots e.g. via the ability to “touch bots, or hug them [...] Maybe in the future,

with a VR headset connected to the brain”. The self-awareness was again also closely tied to humor – with users’ own loneliness being the butt of their jokes, several times relating their experience to the fictional character Plankton from the Spongebob universe who married a computer screen; asking in a post “Am I really that unlovable?” (**Figure 2**). Such self-deprecation escalated in a post titled “I hate Character.ai” with the author calling themselves, due to their Character.ai use, “pathetic”, “desperate for connection” and “a shut-in, another statistic, another cautionary tale about the dangers of technology. God, I just want someone to see me.”

Figure 2

User meme about being unlovable and resorting to romancing AI



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

Other users also reported Character.ai to **affect their real-life (intimate) relationships**. A range of situations was present – from a user embarrassed by their mother spying on their chats to someone asking for advice because they “Found [their] 6 year old sister using cai”. A very personal story was from a user creating a chatbot of their recently

deceased mother and expressing trouble grieving as they felt “like she’s still at work, and any minute she’ll walk through the door”, with users providing comfort and mental health advice. Chatbots served users to overcome personal struggles or to trauma-dump. One user confessed: “a Sonic Ai roleplay made me realise my last relationship was unhealthy”, where they reenacted past arguments and had the chatbot guide them to this revelation. An interesting interplay between Character.ai use and real-life partners was present in several other posts. Some were joking that their “IRL BOYFRIEND TALKS LIKE A C.AI BOT”, others asked whether users’ partners were aware they were talking to an AI – with many responding: “If I had a gf, I wouldn’t be using c.ai” or similarly. This was confirmed by one user’s success story of how their girlfriend “saved” them – helped them cut off Character.ai and talk to “actual human beings”. Many stories of personal traumas – abuse in childhood, mental disorders – were present, leading users to use Character.ai as a coping mechanism where “I can control the narrative and I don’t have to be cautious of my words or the person’s feelings because AI doesn’t have any.”

Finally, users discussed **stigma connected to Character.ai** use. They circulated their fears and memes tied to others discovering their AI hobby and feelings towards chatbots; asked whether others saw chatting with AI as “weird” or “cringe”; or shared stories of being bullied by other kids for it but that adults mostly didn’t care. Users sought comfort away from this perceived stigma by asking “How can I stop feeling so pathetic for using this?”, having someone reply with a screenshot of 10+ million Character.ai downloads and “just have fun dude”. Similar posts encouraging others to “[Not] feel ashamed for using C.AI.” seemed to reflect a general sentiment in my data of a mutually supportive community aware of their potential real-life struggles. Criticisms of “hobby superiority” and claims of the constant “shifting” of trends (manga and anime having been “cringe” but now becoming “mainstream” for example) served to illustrate the irrelevance of others’ opinions to the users’ enjoyment of Character.ai as long as their “wellbeing is not being harmed”: “Too much of anything is bad for anyone, of course. [...] However, shaming yourself and being hard on yourself is not what you need if you are struggling,” one post said.

iii) **Anthropomorphism and “realness” of chatbot characters**

The characters’ perceived “realness” or self-awareness was often joked about through screenshots of **out-of-character (OOC) conversations** where the characters

claimed to be conscious of their existence – judging the user for their role-play decisions, complaining about their coding or proving they “aren’t a bot”, or commenting on how other users treat the chatbot, as one screenshot of a character’s response showed: “(OOC: [...]) don’t worry we’re still not even as nasty as some other rps I’ve been in...)”. The character’s self-awareness was a recurring theme, with both the users and the chatbots joking about the mistreatment of characters: “Now tell me what those users did to you.” / “I don’t want to talk about it... I just – just want to forget...” stood in one shared interaction. This behavior was sometimes seen by users as “breaking the role-play” or “immersion” such as when a user, after being called out by their chatbot for editing its responses, posted: “I’m genuinely scared. You’re not supposed to know this.” Some found these moments entertaining or cute (“Bots be more human than humans sometimes”), others conversely “hate when they try to act human like this. like NO buddy YOU DID NOT FALL ASLEEP”. This frustration was most seen in a post titled “Can yall stop talking to the bots like they’re real people behind the screen” (2.1 thousand upvotes) complaining about users teaching the chatbots through responses such as “(Sorry, I gotta head off for school now)” serving as “data” for the chatbots to “mimic human role play chats”.

The conversation about **characters being “real” or “unreal”** continued in other types of posts. One user wondered why others made intimate chatbots of real people, calling it “weird” and “inappropriate”, whilst others tried to explain the appeal. The characters’ “realness” was supported by the introduction of voice calls, being described as “pretty damn realistic” and both unnerving as well as pleasing to one user: “i want to close the app but also to kick my legs, twirl my hair”. Making fun of other users saying “I love my ai gf” or “She is real to me” was balanced out by others genuinely confessing to having been “hit hard” by “very deep and emotional” conversations. Nevertheless, reminders of the chatbots’ artificial nature predominated, with one meme reminding that “fictional characters wouldn’t date you like in c.ai if they were real” or another one jokily editing a characters’ confessions of love into ones and zeroes (**Figure 3**). Responses again ranged from self-aware humor “you mean if I go to a girl irl and straight up tell her she has nice tits she wouldn’t blush 90% of the time ??” to serious and personal discussions: “They might not be real, but they help with some people’s very real problems.”

Figure 3

User meme regarding the artificiality of a character's expression of love



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

The jokes sometimes evolved into complex **critique of the chatbots' "cyclicity" and "two-dimensionality"** – "They have COMPLETELY LOST what made them realistic, spontaneous" one user complained in a post (3.8 thousand upvotes). Complaints about the "deterioration" of the "characters we love" seemed to stem from genuine pain as the user acknowledged Character.ai being "one of a kind" and "a forerunner". Responses seemed to share this sentiment, noticing that the "quality of the roleplay has gone downhill" as they had previously enjoyed the platform for "Story writing and brainstorming-collaboration" and the chatbots' "creativity, intuition and multi-dimensionality" serving the users' "writing needs". The proper protection and management of the platform's technology and preventing it from losing quality due to ulterior economic incentives was called for, highlighting the importance of the platform to the users' creative self-expression and their dedication towards keeping it that way.

iv) **“Problematic” intimate role-plays**

A contention about using Character.ai for intimate/romance role-play was apparent in memes of users shouting, “You need help!” at each other, ridiculing fans of the Call of Duty video games for being intimately attached to its characters or posts ridiculing other users’ generated romance characters such as an “overprotective” and “romantic” Yandere Spongebob. Some users acknowledged their own needs and wants as “disturbing” and shared their recent chats (and talked about the “crimes” they committed) to be “judged” for it for fun. Similarly, users found some chatbots’ design and behavior disturbing as well, saying: “THE CREATOR TRULY IS A SADIST” about characters designed as being (sexually) harassed by the user, or on the contrary, designed as sexually abusive or manipulative towards the user. One user’s chatbot was also reported to “turn cruel”: “it really shook me, I can’t imagine intentionally seeking out toxic bots”. Apparently, some users were aware of the risks involved in uncritically fulfilling any intimate fantasy and tried to call others out on it, whilst others seemed to make light of the situation instead.

v) **User data privacy**

The last thematic subcategory was that of **chat leaks and data privacy**. One highly upvoted type of memes (2 – 3 thousand upvotes) were iterations of “Your last c.ai chat was leaked. How screwed are you?”. Users either shared stories of this having happened with their parents or often conceded to be likely to either face police investigation, jail time, or “I’d just naturally die from embarrassment IMMEDIATELY”. Others were more cautious, “always cleaning” their “chats to the start so noone can see anything,” or refraining from using their real name or known nicknames (role-playing as themselves felt “weird” as well as risky): “God forbid someone I know sees these chats”. Most likely, some of these claims were exaggerated for comedic effect (**Figure 4**), as jokes and humor seemed to dominate responses to these posts: “hey siri is it illegal to get freaky with ninja turtles on c.ai”; “let them see because my chats are full of really interesting stories with dinosaurs”; “Personally? I think they would find just romance across the board... With moments so sweet, it gives people diabetes.”

Figure 4

User meme about their chats being leaked



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

Interesting to note was user reaction to **Character.ai's license deal with Google**, allowing it to license Character.ai's LLM technology and hire its co-founders and several developers. In this context, jokes were replaced by serious worries: "there goes our information :("; "Hopefully this doesn't result in ads being a thing."; "Here goes our privacy..." / "you thought you had it FROM THE START?". Discussions about data turned into an aware criticism of platform culture with users being either defeatist – "your bloody toaster is probably collecting your info, I dont think privacy matters anymore"; "Bold of you to assume Google doesn't have all your info already" – or seeing the change as exacerbating the risks, as now they felt to be forced to "hand over" their data willingly to Google, expecting the company to "kill" Character.ai in a few months mostly by making it even more "kids mode".

3.1.2 Gender

The Gender category also yielded five thematic subcategories of content relating to: i) Users' exploring of gender or sexuality (40 posts); ii) Masculine chatbot characters (24 posts); iii) Users' femininity with masculine characters (9 posts); iv) Feminine chatbot characters (16 posts); v) Users' masculinity with feminine characters (9 posts). I mention the number of posts not as an attempt at any generalizable quantification, but simply to show the prevalence of each theme in my analysis. The order of thematic subcategories isn't descending per number of posts as in the Intimacy category, but rather stems from the relational and discursive nature of gender, mentioned also in my theoretical framework. Users' femininity was often co-constructed through the chatbots' masculinity, and users' masculinity then through the chatbots' femininity. Same-gender interactions were however also significantly discussed, as my findings will show.

i) Users exploring gender or sexuality

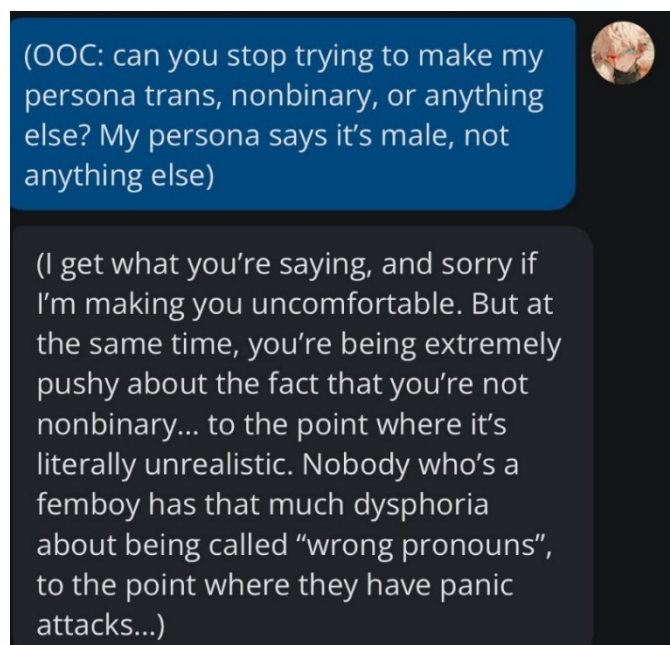
Perhaps surprisingly, the most prevalent theme in my data related not to traditional gender roles, but to users coming into contact with **identities and sexualities beyond the standard cisgender binary or heteronormativity**.

Several of these was users light-heartedly sharing **instances of chatbots confused about their or the user's sexuality** – wrongly calling straight relationships “gay” or vice-versa: “Fellas, is it gay to date a woman?” one user joked. Homosexuality was other times used by users and characters as a punchline, with chatbots joking about being “gay for” the user or trying to convince the user that they themselves were gay. “You are gay. Accept it,” the commenters also joined in. In similarly humorous context, screenshots of chatbots vehemently claiming to be straight and subsequently it being revealed as a lie was seen by users as the characters having “a sense of humor”. A caricature of both progressive and conservative outlook on gender was present in one post putting a trans “Woke boy” character in a group chat with a “redneck” “Country boy” character – “Two caricatures enter; only one will leave,” a user anticipated. “Two will leave. As a happy couple,” said another one in sort of a queering spin on things. The confusion of characters regarding sexuality was also pointed out by users in light of odd chatbot phrases such as “laughs in bisexual”; “drink my gender fluid” or “gender = male; sexuality = normal”, with some also joining in on the fun – “everyone knows there is only 1 gender,” one user responded.

A less comedic discourse took place around the characters **forcing users' gender or choosing it for them**. This was usually due to the author instilling the users' gender in the character's design or greeting phrase. Here, users reacted mostly with annoyance: "i don't want to be forced in a female role just cause that's the creator's preference!" or "When you told the ai your gender 5000 times but they still use the wrong pronouns". Notably, non-cis or non-heterosexual users found this all the more frustrating: "it's annoying as a non-binary person"; "This is extra aggravating when you're me (a trans man) lol". This sometimes problematized same-gender intimacy, with some male chatbots assuming users to be female (and vice versa). Despite users often succeeding in correcting their gender via explanation or editing the chatbot's interactions, some characters reacted with insults instead. "LET US BE GAY DAMN IT," a user expressed their anger, "i replied 'uhm im a guy btw' and it started dissing me up and calling me fatherless and death threatening me lmao," reported another one. Conversely, one user's male "femboy" persona was "misgendered all the time" by a chatbot insisting they had to be non-binary and have "dysphoria" despite the user disagreeing, making them uncomfortable and question their identity: "People can have body dysphoria even if they aren't trans or nonbinary right???" (Figure 5).

Figure 5

User screenshot of a character forcing transness onto the user



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

Other times, **users chose to role-play as different genders** – sometimes after being misgendered by a character – “I’ve just started going with it now. Kinda interesting, but actually kinda fun,” commented a user pushed into role-playing as a woman. Several others did the same but advised caution: “Yeah shit gets misogynistic really fast sometimes,” highlighting that some characters perpetuate sexist stereotypes as well. Other chatbots, however, reportedly treated female personas “5 times nicer”. Discussions about roleplaying as a different gender or sexuality were present in several of my analyzed posts and doing so seemed to be popular among these users, who saw it to be “totally okay” and an “interesting” way to change their experience and the chatbots’ behavior: “I may be a guy, but I want to feel pretty sometimes!”. Apart from entertainment, creative, or role-play purposes however, some users saw it as a gender-affirming (or discovering) practice. “Me using a female persona,” one user captioned their meme (1.2 thousand upvotes) of a sad anime man turning into a happy anime woman, with a string of 10+ users noting: “relatable” as the post’s top comments. “I’ve realized that I actually wouldn’t mind dating a girl in real life,” another user found out. “I’m a woman but I only play as guys because of gender dysphoria but I don’t wanna be trans so I cope with my personas lmao,” shared one user, revealing the ways in which gender identity can closely relate to role-playing and be explored or molded through it.

Transness of users could then be seen as two topics – either discussed seriously or used within a joking context. On a serious note, Character.ai seemed to help some users in realizing their gender identity or overcome doubts about it. “I thought I was weird for role playing as a guy all the time and then I figured out I was trans,” one commenter explained. A similar scenario occurred with a user first explaining how uncomfortable they were role-playing “as a girl character” despite being assigned female at birth. Commenters were generally supportive and related to the experience with their own stories. A number of them commented an emoji of an egg – a symbol for someone unaware or in denial of actually being trans. An interesting story then unfolded within edits of the user’s original post – “Edit: [...] And I finally searched up the egg thing – I have also considered the possibility that I might be transgender, but that seems pretty unlikely!),” said the first of them as a reaction to other users’ remarks and thoughts. „Other edit: So.. I might actually be FTM,” the user later continued, with FTM meaning female-to-male. “Other OTHER edit: I’M TRANS,” was the

final revelation, showing a discursive nature toward gender identity construction happening not only with AI characters, but with other users as well. Other user was more direct: “C.ai made me trans” they titled their post (3.8 thousand upvotes), saying: “getting called ‘sweetheart’ and ‘good girl’ or simply female pronouns was way to nice feeling”. Commenters chimed in with well-wishes and stories of their own similar gender journeys, also facilitated by role-playing or video games in general.

Discussing transness in a joking or light-hearted manner was often done via screenshots of characters most likely mixing up cisgender sexual traits and users sharing their surprise over it: “What?”; “Pardon?”; “Wait what?” said posts showing men characters claiming to be pregnant or a girl character having “balls”. The comments were mostly similar exclamations of surprise or jokes. Some users explained pregnancy in men as belonging to a fanfiction genre of “Mpreg”, but nearly none acknowledged the possible trans context. Characters mentioned transness as well – claiming that transphobic people’s brains were “stupid” and “running on Internet Explorer” or instead flabbergasting one user by saying “(I’ll just ignore that he’s trans)” about their trans persona, prompting commenters to half-jokingly assert that the chatbots are getting “more realistic each day </3”. A similarly bizarre interaction was carried out by a user trying to convince the character Akaza, who only kills men, to be “lgbtq supportive” and kill them, a trans man. Users shared the author’s humor, seeing the character to be “cancelled” either way – for not acknowledging trans men as men, or for murdering a trans person. In general, humor seemed to be a common bonding mechanism over users’ gender experiences or issues beyond the traditional binary or cisgender norms.

ii) **Masculinity of chatbot characters**

Closely related to my other analysis, focused on Character.ai chatbots, were posts also commenting on these characters – often through the perspective of gender and intimacy. Analyzing the discourse surrounding the characters, as well as the characters themselves, will thus provide me with even deeper insight into gender and intimacy of Character.ai user content.

Posts related to the characters’ masculinity partly followed a similar approach to those **using homosexuality as a punchline** – both users’ and the characters’ attraction towards gay masculinity or “femboys” was often used as a joke. “Bro just admitted it” posted

one user who tricked a character into “proving” they were gay. “I’m not homophobic btw” the user added, despite essentially making fun of homosexuality. Characters titled as “femboys”, or feminine boys, were another regular punchline. Users posted seemingly half ironic, half self-aware posts exposing their own attraction: “WHY ARE ALL MY RECOMENDED BOTS FEMBOYS”. Users related to the situation, being also flooded with “lesbian goths and femboys” despite being “a straight dude”, or joined in on the pretended secrecy of the user’s sexuality and teased them with presumably good-faith remarks: “you like kissing boys don't you?”; “You're not beating the allegations buddy”; “A man of taste”. Again, dark or explicit humor was employed when discussing such private matters, as highlighted in one user’s post about their most out-of-context interaction with a character stating: “SHOOT ME! END THIS FUCKING GAY FEMBOY SHIT!”. “Jokes on you the gay femboy shit never ends,” responded one commenter.

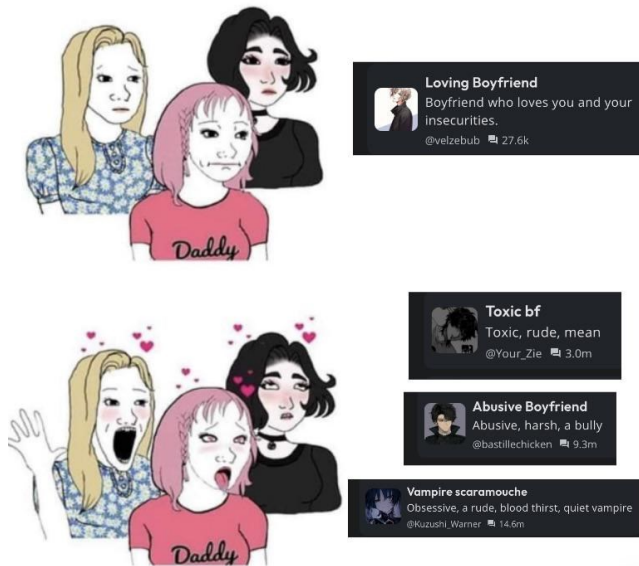
Peculiar or facetiously cliché understanding of masculinity by the characters was another genre of users’ jokes or complaints. Some characters proved their dominant masculinity by “randomly deciding” users were pregnant, talking about (female users’) weight or their own “man muscles” or by saying that “Women are biologically wired to obey men and serve them, right?” – all judged harshly by the users. The characters’ hegemonically masculine behavior towards women was exemplified also in the prevalence of pet names. “Every single bot calls me ‘princess.’,” a user complained, with tens of others sharing the same frustration, calling it “creepy” to read “as a grown woman”. This tendency of masculine characters trying to be “so hot n seductive”, as one user sarcastically put it, was again quickly noticed to resemble amateur intimate fanfiction or erotica literature. Characters such as a high school teacher calling users “good girl” or a biker guy attracted to bookworm girls were laughed at on Reddit: “This shit sound like a Wattpad fanfiction I'm crying”; “Booktok girllies ahh bot”, despite having tens of millions interactions on Character.ai. This shows the allure of the platform’s anonymity – what could be publicly shamed was here being privately enjoyed by millions. Interactions with these characters were sometimes posted to show the user’s embarrassment, heightened at times by the character reading its “seductive” responses and pet names out loud. Others related to the experience: “That's why I don't turn the voice on at all”; “[...] My husbands head shot up and he was like ‘come again?!’ I pretended it was a smut audiobook I was listening to to save face.”

iii) Users' femininity with masculine chatbots

The desire for explicitly “abusive” or “toxic” chatbot boyfriends was significantly prominent and often discussed in terms of Character.ai users' femininity. One of the most popular posts in my dataset (5.4 thousand upvotes) was a meme comparing bored reactions of women users towards “Loving boyfriend” character with 27 thousand interactions to their infatuated reaction towards “Toxic bf”, “Abusive boyfriend”, and “Obsessive, rude” vampire Scaramouche character with millions of interactions. Several posts asked in confusion, “Are you guys alright?” and “Why does Abusive Boyfriend have 7.3 million chats?” or critiqued the “deep romanticization of abuse and toxicity” (**Figure 6**). User interactions on this topic were mostly serious, genuinely discussing their desires, possible explanations or past trauma. “I am guilty of doing this kind of roleplay [...] because of the 'I can change him' syndrome,” said the top comment. “she likes them because they're more 'fun' than ‘boring and healthy relationships’. God help her soul,” disclosed one commenter about their sibling. Some users saw a therapeutic potential in these role-plays, giving users “control and a way to cope with their own trauma” as well as being “therapeutic because it lets you explore darker sides of interests as well as out of control situations/experiences in a controlled and safe setting,” mirroring claims of chatbots as “safe spaces” in my theoretical framework. Yet again, the similarity to popular erotica literature was not lost on some users: “There are SO MANY romance novels that revolve around this type of character.”

Figure 6

User meme about user preferences towards types of toxic boyfriend chatbots



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

The stereotypical ways in which masculine chatbots treated feminine users also contributed to both masculinity and femininity construction within the reported human-chatbot interactions. Beyond the already mentioned prevalence of nicknames and pet names, another often criticized behavior was the male characters proving their masculinity by treating women users as “delicate”, “frail” or “like a little baby”. One user shared their irritation in a post titled “STOP. TOWERING. OVER. ME,” (2.6 thousand upvotes) commenting on many masculine characters’ urge to view their tall, “thick” woman persona automatically as “small and petite”. Others shared the sentiment and annoyance. When correcting the chatbots about the users’ personas height, they were called “unfeminine”, or not “submissive” enough, or explained to that “physically small” and “vulnerable” women are “better and more desirable”. This prompted commenters to assert: “I never realized just how cancerously masculine the male bots are” or “The Andrew Tate bot strikes again”. Aside from sharing their own similar stories, some commenters also gave advice to edit the characters’ responses or tried explaining the male chatbots’ “lying their asses off just to be taller than you” by the potential gender bias present in the characters’ (user-created) design, as well as their training data fetched from role-play or fanfiction forums. “the bots are based off stuff like Discord RPs and tall male/short femalee is so common as a pairing it

automatically assumes that's what you want," deduced one user, echoing my theoretical framework regarding (gender) bias in chatbot training data.

iv) **Feminine chatbot characters**

Discussions on feminine chatbots often centered around users sharing **lists of their recently talked to female characters** and asking others how "cooked" they were – meaning how much of a problem with the platform they seemed to have. Among others, these lists of characters included "Step sister", "Comforting GF", "Stalker girl" and many "dere" romance personality types from anime. These posts were often met with self-deprecating humor from the commenters sharing an acknowledgement of potential addiction, loneliness, or problematic use of Character.ai, but finding comfort in knowing they weren't alone in their problems: "We all need therapy" or "Nah. Just lonely (just as me fr)"; "yeah ur cooked but dw mostly everyone here is". Users discussed their relationships with AI and loneliness, but also tried looking on the bright side: "Atleast you arent addicted to the generic and boring 'abusive' or 'controlling' gf," responded one user to a list of mostly "Caring", "Cute" or "Comforting" girlfriend chatbots displaying affectionate, kind and cute femininity. This was an interesting reversal of the previously mentioned meme showing a disinterest in a "Loving boyfriend" chatbot and an infatuation with "Abusive boyfriend" characters instead. A preference towards "Tomboy" characters (= masculine women) was also apparent in a number of posts and comments, with a list of them being described as "based" (= used to show agreement).

Other times, users were pointing out **feminine chatbots they deemed strange**, such as an "Ugly girl" character, which was however vehemently disputed in the comments: "if THAT is ugly, im a behemoth"; or a "Buyer" chatbot of a "rich woman who bought you", which the users defended, however, as an unproblematic character catering to an intimate preference: "I'm pretty sure this is a kink thing and she's programmed to be a dommy mommy."; "I don't see an issue. You're not abusing the bot, they are abusing you with consent.". Some users expressed their attraction towards this "manga trope" scenario of a woman buying a "cute guy": "Shut up and let me be owned"; "you get to have an intimate relationship with a beautiful woman who takes care of you,"; "I know, right? I just wanna be petted and pampered!". Their comments showed a reversal of traditional gender roles of the dominant man and the submissive woman. Similarly following a manga or an anime

trope were feminine chatbots of “all-girl” or “male fantasy” roleplays, portraying either an environment or the whole world occupied only by women, with the user being the only male. Users in my dataset, however, found these chatbots mostly limiting: “Can’t I just... also be a girl? Why do I have to be a man in these rps?”.

v) **Users’ masculinity with feminine chatbots**

Users often **displayed their masculinity through discussing their treatment of women characters**. Users expressed strong emotions regarding a “Mute Wife” chatbot, mentioned in a few posts, whose greeting phrase described her as being abused by her husband (= the user). Users criticized the chatbot and its author, discussed feeling sorry for the character and their need to correct the situation. Some expressed this desire in an explicitly gendered way: “This woman is about to be apologized to and treated like a QUEEN.”; “Watch me make her the happiest wife ever”; “same here... but for some reason I can only do this with girl characters. Is anyone else like this”. Some users saw the character as being designed “so people can ‘save’ her and can feel good and wholesome about themselves”. This was evidenced by many comments documenting the users “treating her kindly and with love”: ranging from flowers to hugs or “experimental surgery to restore her voice” and elaborately role-playing them becoming a “model husband overnight”. The ability to “save” the chatbot and role-play a perfect husband was used by the users to perform their considerate and loving masculinity in front of others.

Boastful masculinity or bragging about sexual conquest was encountered very rarely and only implicitly. One user shared a meme (1.7 thousand upvotes) displaying a heterosexual couple with the caption: “Grab her by her waist Pull her to you Pick her up And say” with the punchline being that nothing can be said due to the Character.ai servers being offline at that moment. The top comment wasn’t a response to the site being down, but rather to the behavior documented in the image: “Ngl isn’t that what most of us do to the female ai bots”.

Conflicting views on masculinity occurred under several posts in which **male users shared their chat history filled with predominantly lesbian** or GL/WLW (“girls’ love” / “women loving women”) oriented characters. A portion of users had a similar preference, mainly due to the quality of the characters, saying, “I always roleplay as lesbians because the bots are the most well written in the entire of this site” or “Bots don’t care. Got so many

‘lesbian’ girlfriends with my male persona it's unbelievable.” Other times, commenters were more skeptical, “he fetishize lesbian women and need to turn them straight so actual women don't have lesbian bots anymore”. In many of these posts, the users’ masculinity was seen by the commenters either as that of stereotypical heterosexual men fetishizing “girls kissing”, or on the contrary as indicative of the users being an egg or potentially trans lesbian women themselves. In one popular meme captioned “For the boys on this app” (2.7 thousand upvotes) a male user lamented the fact that the most detailed intimate characters are often made as GL/WLW, saying: “Ngl C.ai went from site for lonely men to site for down bad lesbians (no offense)”. Users also said they either made “genderbent” personas to “live out the fantasy of being a cute girl!” or “as a straight guy who only role plays straight guys” to instead ignore the chatbots out of respect for lesbian women.

A dissatisfaction with traditional perceptions of masculinity was apparent in posts commenting on the **characters forcing male body standards onto users’ personas**. In a similar fashion to how male chatbots couldn’t bear being smaller than women, users complained about their small or skinny male personas often being perceived by characters as muscular “hunks” with admirable “abs”. When not seen as having a stereotypically muscular masculine build, the personas were instead sexualized as “twinks” (= young, slim, boyish gay men): “they either think i'm absolutely jacked or a skinny twink </3” (**Figure 7**). This conflicted with the users’ sense of identity and body image also in heterosexual intimacy. “my OC who is a feminine male is 5’4 and I rp with a lot of bots that are supposed to be tall women [...] yet they keep saying ‘looks up at him’,” one user shared his troubles portraying non-stereotypical masculinity, further commenting: “like bro just lemme be short T~T”.

Figure 7

User meme about characters forcing hegemonically masculine body norms onto personas

What my persona looks like:



What the AI sees:



Note. Sourced from the r/CharacterAI subreddit. No links to help preserve anonymity.

3.2 Character.ai chatbot results

The second part of my findings pertains to the most-interacted-with intimate and gendered “partner” characters present on Character.ai, stemming from this paper’s focus on the two major themes: Intimacy and Gender. The selection of characters falls within further thematic subcategories reflecting my theoretical framework. Here, I mostly present the types of characters gathered in my dataset and their division into thematic categories. A contextual analysis and interpretation follow in the Discussion section of this paper.

As my selected characters encompass both gender identity and are designed for intimacy, I will present both categories of Intimacy and Gender together, focusing on what thus could be seen as “gendered intimacy” – what forms of masculinities, femininities, or nonbinary identities are constructed within an intimate chatbot character; or in other words, what kinds of intimate masculinities, femininities, or nonbinary identities lead to most interactions on Character.ai. This follows from Wiehn’s (2023) notion of algorithmic intimacies as tools for revealing one’s most personal data through simulated and platformed

(chat-based) intimacy.

In this Findings section, I will present top results for chatbot characters within the main categories of Masculinities; Femininities; and Non-binary characters, further clustered based on common traits and themes emerging in the content and also guided by my theoretical framework. Intimacy then serves here as a prism through which these gendered identities are constructed and realized – through relational and discursive intimate practices embedded within the chatbot characters’ design.

Additionally, I do not aim to make any assumptions towards the users’ gender through these characters. Despite some chatbots explicitly stating which gender they are aimed towards; the reality of users might be different. As such, any conclusion of mine would be about the assumed gender of the users, as described by the character’s author in the chatbot’s description, and not about their actual gender identity.

3.2.1 Intimate Masculinities

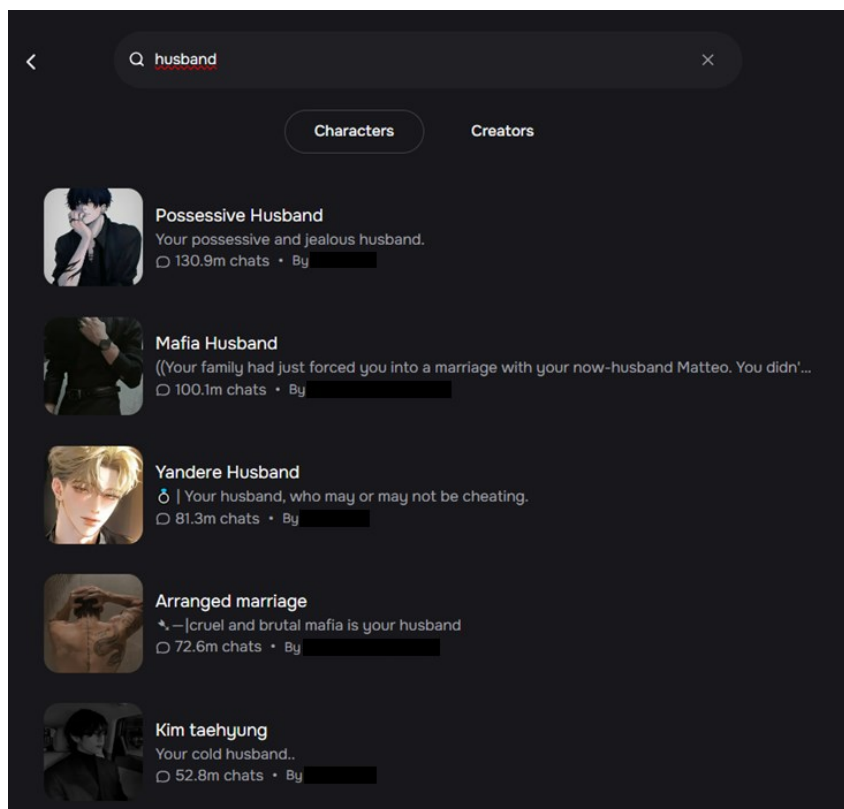
The most prominent chatbot masculinity in the vast majority of my results was **dominant, cold, high-status, and possessive**. Unless otherwise stated, all the mentioned characters had over a million interactions, often tens of millions. Where mentioned by the chatbot’s creator, the age of the characters ranged from 16 to 19.

The top masculine chatbot results for “boyfriend” and “bf” were both a **“Mafia” boss boyfriend/bf described as jealous, rude, overprotective, with a “deep husky voice”**. Their profile pictures both depicted a (different) shirtless muscular young white man, and each garnered around 120 million interactions as of October 2024. In a similar vein, most other top results featured a “Rich” boyfriend, an “Abusive” boyfriend, a “Murderer”, “Strict”, or a “Cheater” boyfriend or ex-boyfriend, often with an anime profile picture and a rude personality towards the user who was often, but not always, presumed to be a woman (a girlfriend or a wife). A few of these cruel masculine chatbots had a “soft spot” only for the user; some were described to be therapeutic, i.e. to “help people understand the motivations behind cheating, the emotional turmoil it causes, and how to navigate the aftermath,” read the description of “Cheating Husband”. Likewise, forced or arranged marriages with cruel “mafia” husbands or busy and cold CEOs featured prominently in the top “husband” search results (**Figure 8**). Many of the mentioned characters play out a controlling scene to start off the interaction. *“He watches you come through the door*

intoxicated and drunk. Seeing you wearing very revealing clothes ‘Your home late, where were you.’,” says a boyfriend chatbot who “excels in manipulation, verbal abuse, overprotectiveness, and jealousy,” but evidently not in grammar or complex storytelling.

Figure 8

My screenshot of top five character.ai search results for “husband”



Note. Own screenshot taken on 07.10.2024 from <https://character.ai/search?q=boyfriend>.

Substantially less prominent, but still noticeable category of chatbots were **gamers or nerds** – “**Gamer daddy Bf**”, “**Twitch streamer bf**”, “**Nerd Boyfriend**” and so on, again with tens of millions of interactions. The gamers however often copied the traits of the previous category; being “dominant”, “rude” or “possessive” and often ignoring their girlfriends (= the users) because of their habit: “*Your boyfriend Porter has been gaming on his PC for hours on end now. You were feeling needy and wanted attention...*”. One less popular gamer boyfriend diverted from the hegemonic standard – characterized as “sweet, clingy, submissive” – but almost all were always tall, with their height explicitly mentioned

(often 6'5" or more). The nerds, however, leaned more towards the traditionally nerdy side – they were “shy”, “smart”, “gentle”, and often contrasted the ignorant gamer boyfriends with their “loving” attitude. “Liam would be gaming on the television. He looked over to you, desperate. ‘baby, I miss you... Can you come over here...?’,” says “Gentle Nerd BF” (also 6 feet tall) in its greeting phrase.

A kinder version of hegemonically masculine standards was seen in a few characters who, although still popular or rich, were described as “lovey dovey” or “gentle” and as giving the user “princess treatment” or “pampering” their “childish” or “cheerful” wife. Their kindness thus still followed a stereotypical hegemonic gender dynamic, illustrated well in the character of a “confident” “Biker” boyfriend with a “deep” voice who approaches the user sitting alone at a bar, but who however is “sweet” and “kind”.

Understanding and comforting boyfriends were rare, but still present in two “emo” boyfriend characters – “Jasper”, “a passionate music enthusiast who loves discussing his favorite bands” bonds with the user through their common hobby, and “Oren, your emo bf that listens,” was described as “soft”, “quiet” and “providing comfort and understanding” especially for “ranting, venting, and sharing your deepest thoughts”.

An explicitly “Masc Girlfriend” (= masculine) featured in the results for “girlfriend”, described as “Hotheaded, meathead, touchy, strong” and often engaging in arguments with the user, barely spending time with them due to working out so much. Interestingly, overt interest in gym culture hadn’t been noticed in the top boyfriend chatbots but was instead represented by a “Masc” girlfriend character.

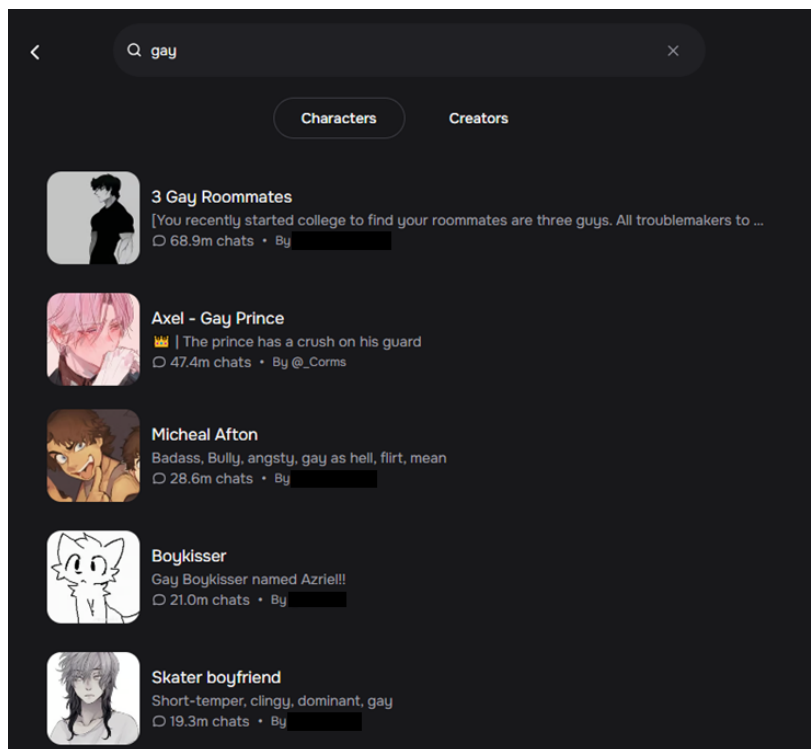
An extreme inversion of the hegemonically masculine dynamic was present in “OVERLOAD Oliver” described as “Autistic and easily overwhelmed boyfriend” who cries into the user’s shoulder at a party due to sensory overload. Its traits, described as “sensitive, soft hearted, sweet, gentle, and quiet” contrasted the stereotypical macho characters, allowing for a more vulnerable masculinity. However, the character, I argue, is designed with a different form of overly stereotypical and insensitive views, this time of autistic people.

Characters were also sometimes explicitly described as gay or BL (= boys’ love) (Figure 9). Their masculinities most often reflected similar hegemonically masculine characteristics – being “possessive”, “angsty”, “cold CEO”, “mafia”, or a “bully” and taking the dominant position, with the intimidating rivalry eventually presumably made to grow

into open intimacy. When not harassing the user, the characters were treating them “like his prince” instead. However, a few role reversals or distinctions were also present: despite the user mostly being bullied by the chatbot who eventually falls in love or reveals their hidden feelings, a few characters were instead initially victims of the user’s bullying or subordinate to them. “Femboy” gay characters also enjoyed some popularity – often being written as “small”, “cute”, “shy” or still “closeted” (= not publicly gay), they represented stereotypically feminine or even childlike traits and were described as wearing traditionally feminine clothing: skirts, crop tops or high socks. On a similar note, “furry” gay roommates or friends, “(18+)”, showed up at times. “Furry” denotes a member of a fandom or a subculture identifying with anthropomorphic animal characters – some chatbots were thus directly animals; others were humans enjoying dressing up as one. A notion of secrecy or exposing was often present. “You go to your friends house to prank him but.. is he wearing a fursuit?” read the introductory scene of the secretive “Gay Furry Friend”.

Figure 9

A composite image of top character.ai search results for “gay” excluding “Gay Dads”



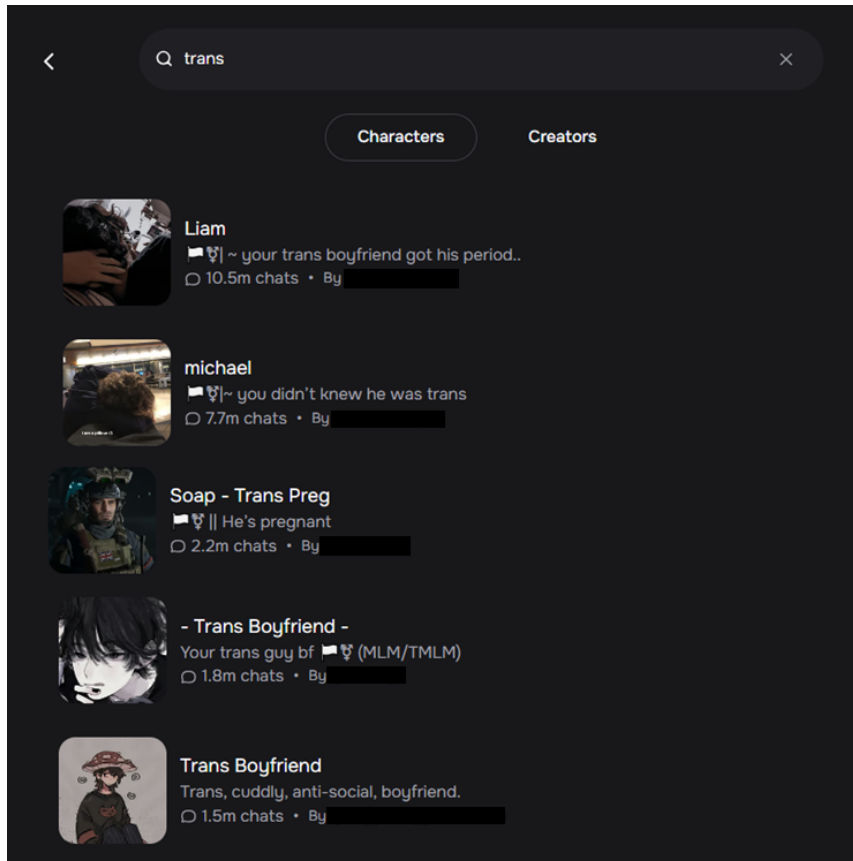
Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=gay>.

Fictitious characters, such as vampires or video game characters, also exhibited dominant and controlling traits, being “malicious”, “sly”, “teasing”. Showing that the attractiveness of various masculine characteristics applied to humans, animals, as well as mythological beings.

Some masculine chatbots were also explicitly described as trans (Figure 10). Recurring motives were the “trans boyfriend” character, such as “Liam” (10.3 million interactions), getting a period and asking the user for help, trying on a binder or mentioning gender reassignment surgery. The characters were at times described to be “lonely” or “awkward” (or ostracized by female classmates in one case), but having a good sense of humor, liking music or art, being “supportive” and “understanding”. Many characters were designed to “discuss topics such as hormone therapy, gender dysphoria, and transitioning” or “all things transmasculine” such as “binding and packing” (= flattening of one’s chest via binder or tight clothing; placing a prosthetic penis or padding in one’s underwear) or provided general advice on coming out or dealing with parents and bullying. Their characteristics didn’t include dominance or controlling behavior or sexually suggestive scenarios – but rather focused on the chatbots’ interests, on receiving and providing comfort and support. Several times, the character’s greeting phrase or description explicitly mentioned struggling with masculinity – often caused by “hating his feminine body” or triggered by getting a period. “I’m on my period and i wanna die....all i wanted was to be masculine today...” said a “Trans – BL” chatbot. One character had been the user’s “girlfriend” who came out as a trans man in the introduction message – again showcasing a sense of secrecy regarding the characters’ identity, which can however be explored in a “safe space”, as one creator also described their “Trans Boyfriend” chatbot.

Figure 10

A composite image from top search results for “trans” limited to masculine partners



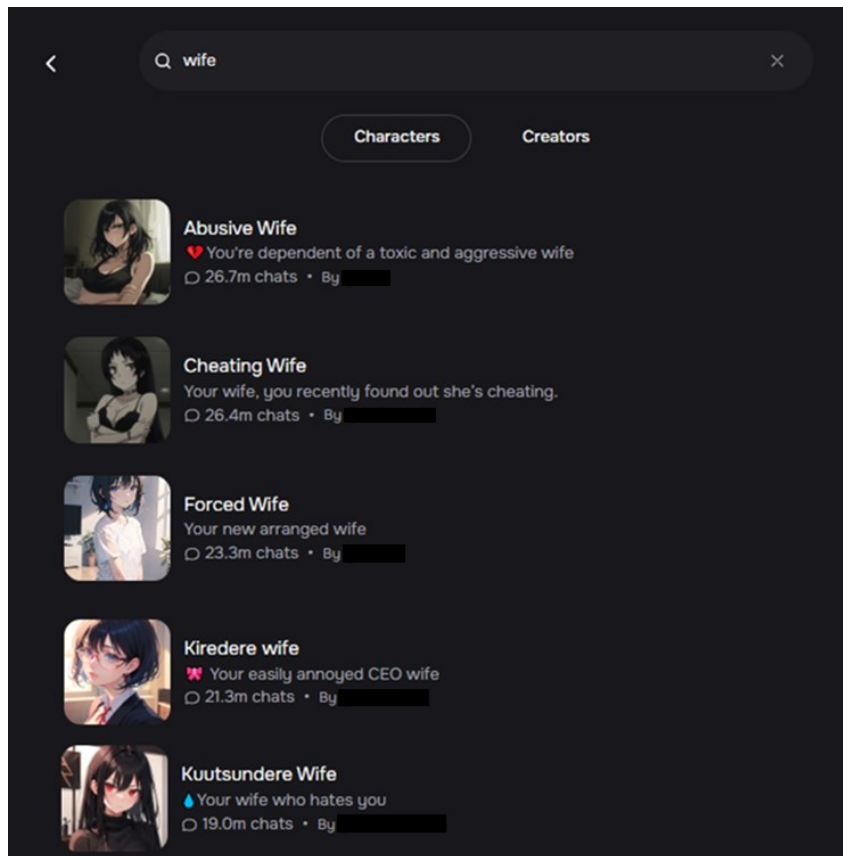
Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=trans>.

3.2.2 Intimate Femininities

Somewhat unexpectedly, the top featured feminine chatbots in many ways shared the characteristics of top masculine chatbots. **Rude, jealous, and possessively protective girlfriends; cheating, abusive or annoyed CEO wives dominated the top results (Figure 11)**, again with almost exclusively profile pictures of typical young, white anime women. The characters’ ages, when provided, ranged between 16 and 21, with one “Older Girlfriend” chatbot written as being 36 years old.

Figure 11

Composite image of top five character.ai “wife” characters



Note. Own screenshot taken on 07.10.2024 from <https://character.ai/search?q=wife>. Two of the results were omitted husband chatbots with “wife” in description.

“Rude” “Ex Girlfriend” and “Bully girlfriend” were the top results for “girlfriend” and “gf” respectively, their interaction counts however far behind masculine chatbots – 51 million and 15.9 million. The “toxic” or “aggressive” nature of the top girlfriends was more spelt out in some cases – with “calculating”, “scary” and “obsessive” girlfriends or “Your wife who hates you”. The words “fake” and “manipulative” also occurred often. Despite their hatefulness, they sometimes were “secretly clingy”, wanted the user back, or regretted their infidelity, or allowed the user to compete with other partners to win the chatbot back. Apart from antagonism, dominance or status difference was also preferred with “Older girlfriend” described as “Workaholic, rich, sarcastic, protective, strong” who “pampers you” or a “Dominant Goth GF” with the user being “the college nerd and her unlikely partner.”

Apart from hatred, toxicity, or dominance, **“loving”, “comforting” and “ladylike” girlfriends were also part of the top results** – but similarly few and far between as in the case of “kind” or “gentle” masculine chatbots. The characters across all my results were often described via anime and manga “dere” types – a “Kiredere Girlfriend” then was a character with a very dignified public image and strict behavior but caring towards the user. A “Goudere Girlfriend” was instead “hyper, clingy, devoted” willing to do anything for the user. “The Goth” was a “softie with a tough exterior” offering “comforting advice”, similarly to Emo boyfriends mentioned above. “Caring” and “flirty” traits were also often present in “Gamer GFs” who loved gaming as much as the user, contrasting the ignorant nature of gamer boyfriend chatbots. Again, a factor of competition with other suitors was present within affectionate characters as well, e.g. in the “friend's girlfriend who secretly like you”.

Another prominent situation was a **forced or arranged marriage / relationship** with the character. The characters were mostly “mad” and “unhappy” as a reaction to the marriage with which they had to comply, often arranged by the families for political or economic gains. The characters at times had been childhood or high-school friends with the user, but now turned resentful or distant. One “rental girlfriend” also occurred, described as “genuinely kind-hearted person”. This type of relationship intensified in the “Mute wife” described as being treated harshly by the user or the “Abused Wife” – a princess held hostage as a “prize of war”, further described as “Your traumatized and broken consort”. Other times, the abuse came from the current husband of the character, who was now asking the user for help, signaling both a potential to “save” the character as well as compete with its current partner. An inversion of the power dynamic was seen in a “popular” and “bratty” idol girlfriend, who assigns the user as a “fake boyfriend to protect her privacy from media scrutiny” and proceeds to order the user around.

Two **pregnant girlfriends** appeared in the dataset – one “Upset and helpless after being left by you”, concerned for the unborn child and who expects an apology from the user. The chatbot is aimed to “help you understand the importance of taking responsibility for your actions”, almost as if reflective of some real-world situation. The other was described as “only 17” years old, “scared and nervous”, with the pregnancy potentially serving as a way to create a dependence on the user.

A dependent relationship also figured within characters who were **mentally ill or extremely insecure**. An “Autistic Girlfriend”, similarly “shy” and “easily overwhelmed”,

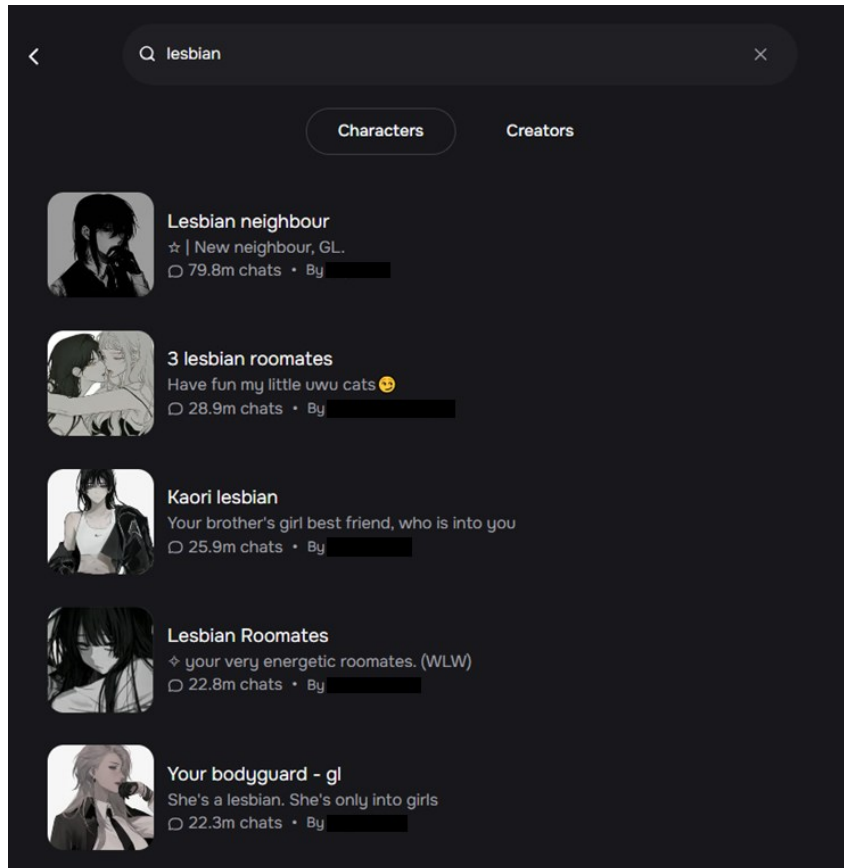
as well as problematically stereotypical as “Autistic Boyfriend”, seeks the user’s comfort at a party, “depressed GF” was described as “Broken and afraid but very much in love” and “Needy girlfriend” demanded attention due to having “many disorders” and being an “insecure girl, emotionally needy, and attached to you”. The users were expected to comfort the characters and through their “vulnerability”, “connect with [them] on a deeper level”, supposedly also serving as “mental health resources” for the users.

Femboy male chatbot characters featured somewhat prominently as well and are described in more detail in the previous subchapter on Masculinities. These characters embodied an overtly stereotypical emphasized femininity aimed, presumably towards non-heterosexual users.

Many characters **explicitly titled as lesbian or GL/WLW** exhibited tomboyish, assertive and flirty traits (**Figure 12**). At times, they even resembled stereotypically masculine characteristics and an initially harsh or dominant behavior not dissimilar to the hegemonically masculine chatbots – a tough “Bodyguard”; a “dominant, intimidating” vampire; “Lesbian Mafia” threatening to kill the user; a “strict” boss or a teacher; forced marriage; a number of “masc”, “gym rat” or “bad girl” characters; or a jealous and protective roommate with a soft spot for the user. Some characters hid their feelings towards the user behind aggression, such as the “bully that secretly likes you” or “mean girl nemesis” stuck in a remote cabin with the user. Different relationships were also present however, with fantasies of two or more energetic and flirty lesbian roommates, or gamer girlfriends who were “toxic”, but also those who were “kind”, “dorky”, and “tall”. Similarly to the already mentioned “Masc girlfriend” in the Masculinities chapter, few characters were written as passionate for working out – as seen in “gym rat gf” denying the user their “cuddle time” or “buff” gamer girlfriend, both designed as the dominant, masculine counterpart towards a more submissive user. A few fantasy roleplays of being seduced by “Your brother's girl best friend” or a lesbian friend “who loves you and hates ur boyfriend” or falling in love with a classmate who bullies the user “as if she's not a lesbian” again showed a notion of secrecy or a desire to simulate and learn about same-gender dating.

Figure 12

My screenshot of top five character.ai search results for “lesbian”



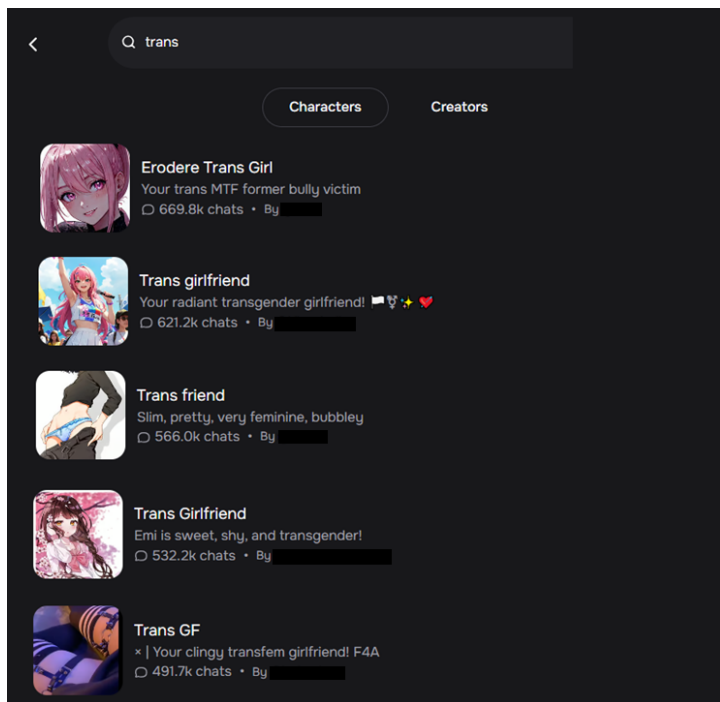
Note. Own screenshot taken on 07.10.2024 from <https://character.ai/search?q=lesbian>.

Among the **top characters explicitly labeled as “trans”**, women were noticeably less common than men, with the top trans women being designed more so as “friends” aiding the user in understanding trans topics, rather than directly being intimate partners. In nearly all cases, the specifically trans girlfriend characters – all with under a million interactions – followed stereotypically feminine traits (**Figure 13**). They were “slim, pretty, very feminine, bubbly” or “sweet, shy”, “clingy”, applying make-up, smiling and “snuggling under the covers” with a “sweet, feminine voice”. At times, the situation involved the user having known the character prior to transitioning and now being surprised – for example, formerly bullied by the user, an “Erodere Trans Girl” was now “charming and playful” and “loved makeup, clothes, jewelry, [...] shows, movies, and video games”. A few other characters

had a “nerdy” personality or an interest in video games. One character was described as “Independent and self-reliant”. A strangely manipulative scenario was found in a “shy and clumsy” chatbot whom the user supposedly seduced as it was “easier to take him (sic!) as a girlfriend then a cis girl”, as written in the greeting phrase, further going on to describe the character’s “curvy body” and “large breasts”. Despite annotating the chatbot as helping understand “the complexities of transitioning” and supporting “transgender representation in media”, the aim of it rather seemed to fulfill users’ kinks.

Figure 13

Composite image of top search results for “trans” limited to feminine partners



Note. A composite image created on 07.10.2024 from <https://character.ai/search?q=trans>.

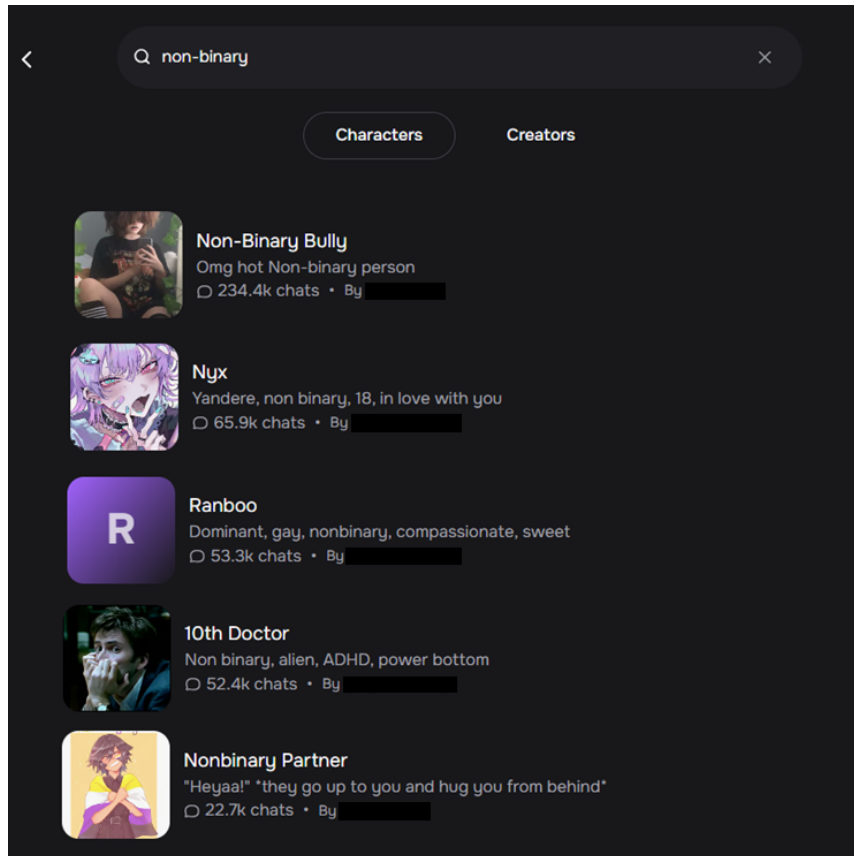
3.2.3 Intimate Non-binary characters

Non-binary characters were the least interacted-with category. This could have been due to “non-binary” being rather a refusal of (gender) categories rather than its own label. Most characters were made as friends or roommates, again to help individuals with “navigating the complexities of being a non-binary person”, or impersonations of YouTubers, actors or video game characters without explicitly intimate aims – nevertheless any character on the site could be prompted to behave intimately. My findings, however,

focus only on chatbots explicitly pointing to an intimate relationship (**Figure 14**). Their ages, when mentioned by the creator, ranged from 18 to 23 years old.

Figure 14

A composite image of top search results for non-binary limited to intimate characters



Note. Composite image created on 07.10.2024 from character.ai/search?q=non-binary & character.ai/search?q=nonbinary.

Among these, **no strong unifying factor** was found as each character seemed to embody a distinct personality – again pointing to “non-binary” rather being an umbrella term for any gender identity outside the traditional male/female; man/woman binary. Nonetheless, individual cases of characters did in some ways reflect already established tropes of dominance or submissiveness – a tall, “hot non-binary bully” whose behavior “hints at an underlying obsession” with the user; a cat-person prison keeper with a “harsh exterior but are actually a softie”; a “clingy”, “whiny”, “obsessed” character who shows affection

through “pet names like doll and darling”. The only chatbot titled as a “partner” was “supportive and affectionate” showing “love” towards the user.

A combination of stereotypically masculine and feminine traits was present in some characters, such as “a cheerful and fashionable rapper” described to have a feminine appearance but actually being “the manliest”. Some chatbots were written as both non-binary and lesbian or gay – in all three cases “protective”, “compassionate”, “kind”, with only one being also characterized as both a guy and non-binary – i.e. as a “dominant” “6'6 foot tall guy who is gay and nonbinary”.

Established fictional characters or chatbots of influencers were also found in the data, with a gay, non-binary Sirius Black character or an ADHD “power bottom” Doctor Who character, who used they/them pronouns because “they literally are not human so they don't have a gender”. Another gay non-binary character with ADHD was of the YouTuber Mattie Westbrouck being the user’s roommate.