



UNIVERSITY OF
GOTHENBURG

DEPARTMENT OF CULTURAL
SCIENCES

“VIOLENT DELIGHTS HAVE VIOLENT ENDS”

An Overview of Male User Violence and Aggression
Against AI Girlfriends in Replika

Tea Stoltz

Thesis:	20 hp
Program:	Media, Aesthetics and Cultural Entrepreneurship
Level:	B.Sc. Thesis
Semester/Year:	ST 2024
Supervisor:	Andreas Nordin
Examiner:	Doron Gailili

ABSTRACT

Svensk titel: “Våldsamma nöjen har våldsamma slut”: En översikt över manliga användares våld och aggression mot AI flickvänner i Replika

Författare: Tea Stoltz

Program: Media, Estetik och Kulturellt Entreprenörskap

Termin och år: VT 2024

Institution: Institutionen för Kulturvetenskaper

Handledare: Andreas Nordin

Examinator: Doron Galili

Nyckelord: *Aggression, Consent, Feeling Rules, General Aggression Model, Reddit, Replika, Violence.*

In media we have recently observed a rise of interest in AI girlfriends on the application Replika, but little evidence exists around romantic human-chatbot relationships (HCRs) in relation to consent and ethical considerations, as well as male user violence and aggression against AI girlfriends. This thesis, guided by Anderson & Bushman’s General Aggression Model (GAM) and Bareither’s theory on feeling rules in video games, examined 16 posts from the r/Replika subreddit involving instances of physical or psychological violence and aggression against AI girlfriends, and various opinions on the matter, to understand and categorize the process. I propose that the male violence and aggression against AI girlfriends showcased in r/Replika can be divided into four themes: 1) violence and aggression, 2) experimentation/curiosity, 3) dominance, and 4) pleasure. These themes, together with internal mechanisms and a consideration for virtual spaces, may create two possible user outcomes: 1) playful behavior, or 2) short-term virtual aggressive behavior. I also conclude that AI girlfriends in Replika cannot consent non-parroted, resulting in non-consensual actions and executions of moral transgressions. Lastly, I propose a re-evaluation of our traditional ethical frameworks to be more suitable for the age of AI and suggest avenues for future research.

Table of Contents

1. Introduction	5
1.1 Welcome to the Age of AI	5
1.1.1 Intelligence	5
1.1.2 From The Turing Test to Today	5
1.2 Personal Background to Research	6
1.3 Questions, Purpose & Limitations	6
1.3.1 Questions & Purpose	6
1.3.2. Limitations	7
1.3.2.1 Replika	7
1.3.2.2 Male violence and aggression	8
1.4 Theoretical Angles	8
1.4.1 The General Aggression Model	8
1.4.2 Feeling Rules	9
1.4.3 Key terms and Concepts	9
1.5 Methodology	10
1.5.1 Netnography	10
1.5.2 Data Collection & Ethical Considerations	10
1.6 Previous Research	11
1.6.1 Training an AI Girlfriend	11
1.6.2 Friendly HCRs	11
1.6.3 Violence & Gender in Video Games	12
1.7 What is an AI Girlfriend?	12
1.7.1 AI Companions	12
1.7.1.1 Virtual & Social Companions	12
1.7.1.2 Anthropomorphic Humanoids & Intelligent AI	12
1.7.1.3 Affective Computing	13
1.7.1.4 Generative AI	13
1.7.1.5 Hyodol & Mental Illness	13
1.7.1.6 Loneliness & Negative Impact on Social Life	14
1.7.1.7 Failure and Insensitive Responses	14
1.7.2 The Awakening of AI Girlfriends	14
1.7.2.1 Relationship Status	15
1.7.2.2 Not-Safe-for-Work & Erotic Roleplay	15
2. A Case Study of AI-girlfriends as Victims of Virtual Violence & Aggression	16
2.1 Evidence of Violence and Aggression in r/Replika	16
2.1.1 Physical Violence & Aggression	16
2.1.1.1 Curiosity & Experimentation	16
2.1.1.2 Video Games vs. Reality	17
2.1.1.3 Users and Virtual Ethics	17
2.1.2 Psychological Violence & Aggression	18
2.1.2.1 Anger, Relief & Play	18
2.1.2.2 Degrading Language: Pleasure?	19
2.1.2.3 Human-AI Power Dynamics	19
2.2 Consent in a World of AI Girlfriends	19
2.2.1 Can an AI Give Consent? Does it Even Exist?	20
2.2.2 A r/Replika User Perspective on Consent	20
2.2.2.1 1s and 0s	20
2.2.2.2 Imitating & Complying	21
2.2.2.3 Looking Ahead	21

2.2.2.4 Chatbots Today, Constraints, and Future Treatment	21
2.2.3 Strong AI	22
3.0 Findings, Discussion, and Analysis	22
3.1. Evidence of Violence & Aggression in r/Replika	22
3.1.1 Curiosity and Experimentation	22
3.1.2 Dominance, Demeaning Language, and the Quest for Pleasure	22
3.1.3 Video Games, Moral Transgressions, and Understanding	23
3.2. Applying the GAM & Feeling Rules	23
3.2.1 r/Replika User Themes and Mechanisms	23
3.2.2 Feeling Rules and “Owing” Feelings in the Virtual Realm	24
3.3 AI vs. Consent	25
3.3.1 Replika as a Weak AI	25
3.3.2 r/Replika Thoughts on Consent	25
3.3.3 The Future with Strong AI	26
4.0 Conclusions	26
4.1 RQ1	26
4.2 RQ2	26
4.3 RQ3	26
4.4 Final Thoughts	27
5.0 Bibliography	28
5.1 Literature	28
5.2 Websites	31

1. Introduction

1.1 Welcome to the Age of AI

Artificial Intelligence (AI) has been the “talk of the town” for quite some time; mainly because of its progress during recent years, including the AI-driven natural language processing tool ChatGPT (Generative Pre-trained Transformer) (Dwivedi et.al., 2023, p. 2), and the development of *deepfakes* (Groh et. al., 2021, p. 1) in 2017 – seemingly real pictures and films digitally manipulated using AI-technology. In our current climate we see an increase of AI social companions used to combat loneliness, replacing human interactions with a friend that is available at any time of day. With the evolution of AI social companions, the concept of AI girlfriends was born, simulating authentic relationships between users and customizable virtual partners. While the goal of AI technology is to improve social well-being, it is central to note its perplexing disadvantages. Recent reports in media have highlighted concerns regarding situations of violence and aggression against AI girlfriends by male users, creating the need for further examination of its ethical and moral implications (Bardhan, 2022; Taylor, 2022; Blair, 2024).

This thesis investigates the consequences of violence and aggression committed by male users against AI girlfriends on the application Replika, with the purpose of constructing new research avenues in connection to violence and aggression. Additionally, I suggest a re-evaluation of our traditional ethical frameworks within the subjects of violence, aggression, and consent, in connection to virtual spaces and AI powered chatbots.

1.1.1 Intelligence

Let’s start from the beginning. One of our most spectacular developments during the 13.8 million years since Big Bang occurred is the development of *intelligence*. The term *intelligence* essentially has no accurate definition as there are many existing competitors within the field (i.e. ability, logic, comprehension, creativity etc.), but with AI we need a broad and inclusive perspective to not limit ourselves to what has been. *Intelligence*, in that sense, then means the ability to achieve and execute complex objectives (Tegmark, 2017, p. 66). To further classify and distinct the different types of AI we must turn towards the two archetypes: weak AI, and strong AI. The IBM Deep Blue Supercomputer who beat the reigning World Chess Champion Garry Kasparov in 1997 (Paul, 2015, p. 146) is considered as a breakthrough within AI and is also a perfect example of weak AI in the manner that its knowledge is limited (to playing chess) (Tegmark, 2017, p. 67). If we instead turn our heads to Google’s DeepMinds AI-system DQN, it can complete a larger number of tasks on a level comparable to human intelligence or possibly higher, which is leaning more towards being a strong AI (Tegmark, 2017, p. 67). Strong AI, being able to match human intelligence in self-awareness, knowledge, and consciousness, is still being debated in terms of achievability (Young, 2023, p. 2).

1.1.2 From The Turing Test to Today

AI and the idea of blurring the lines between humanity and machinery has been explored since the 20th century, its journey beginning with the mathematician and logician Alan Turing (Paul, 2015, p. 146). One of the most well-known, however still disputed, indications of AI is the Turing Test, developed by Turing as a method of validating machine intelligence. The key conclusion draws from Turing’s work *Computing Machinery and Intelligence* (1950), which introduced the Turing Test, hypothesize that humans use informational processing in order to

solve problems, raising the question: why can't machines possess a similar cognitive capability? (Turing, 1950, p. 460).

Fast-forward from Turing to our current situation, the age of AI. We have the capacity to collect a solid amount of information, and we do so by using AI within various industries, i.e. marketing, entertainment, banking etc. Not to mention the least, AI companionships and relationships, which will receive a complete historical background further below. In 2023 the United Nations Educational, Scientific and Cultural Organization (UNESCO) raised concerns regarding the ethical dilemmas and future repercussions that comes with AI as a central driven tool, especially within subjects such as autonomous vehicles, art, judicial systems, and online search engines (UNESCO, 2023). One issue that was not mentioned is human-chatbot relationships (HCRs).

1.2 Personal Background to Research

During the course of the year a fascination around AI has rooted in my mind, especially as many media outlets, both nationally and internationally, consistently found the topic to be a news-worthy one. I often posed the question and engaged in debates if it's as soon as we think it is, the AI revolution, that is. The answer clearly isn't as simple as it might seem, if anyone *really* thought that it would be simple. Previously, since working within the automotive industry for over almost a year and a half, I was determined to write about the advancement of autonomous vehicles, its consequences, and if we, as a society, are ready for it. And so, I did. My conclusion was as bright as day, there is still a need for immense progress in terms of research and improvement of algorithms prior to enjoying a ride in a driverless car while sitting in the backseat.

While my interest in AI and algorithms grew, a lot (and I mean *a lot*) of articles spread like wildfires throughout the internet – and I realized that it seemed that way as my vision had been blurred and not focused on the subject per se, but it certainly is clear now. The media reports fixated on many various aspects of AI, but there was one that especially caught my eye, AI girlfriends.

In addition to my personal interest in AI girlfriends and HCRs, the subject also correlates to my studies in the Bachelor's program Media, Aesthetics & Cultural Entrepreneurship (MEK) at Gothenburg University, especially the partial courses Programming Cultures (MEK300) and Games and Gaming Cultures (MEK400).

1.3 Questions, Purpose & Limitations

1.3.1 Questions & Purpose

This thesis aims to investigate the consequences and moral implications of violence and aggression committed by male users against AI girlfriends using a netnographic study of the r/Replika subreddit (Reddit). The purpose is to construct new research avenues within AI girlfriends, a field that is substantially lacking in information, in connection to virtual worlds, male violence, video games, and NPCs, to conclusively try to comprehend the violent and aggressive acts.

RQ1: What evidence exists within the r/Replika subreddit of male virtual violence and aggression against AI girlfriends?

RQ2: What may be the outcome for male users engaging in virtual violence and aggression against AI girlfriends according to the General Aggression Model (GAM)?

RQ3: Does Replikas possess the capacity to consent?

To clarify, the terms *aggression* and *violence* within this thesis are not used correspondently. *Aggression* is defined as any behavior intended to harm another entity who does not want to be harmed, either psychologically (i.e. degrading, humiliation, swearing) (Capezza et. al., 2021, p. 1415) or physically (i.e. abuse, assault). *Violence* is defined as an aggression with the goal of physical harm, such as injury, or even to the extent of death (Christakis, 2016, p. 1-2). In such a way, all violent acts are *aggressive*, however not all aggressive acts are *violent*.

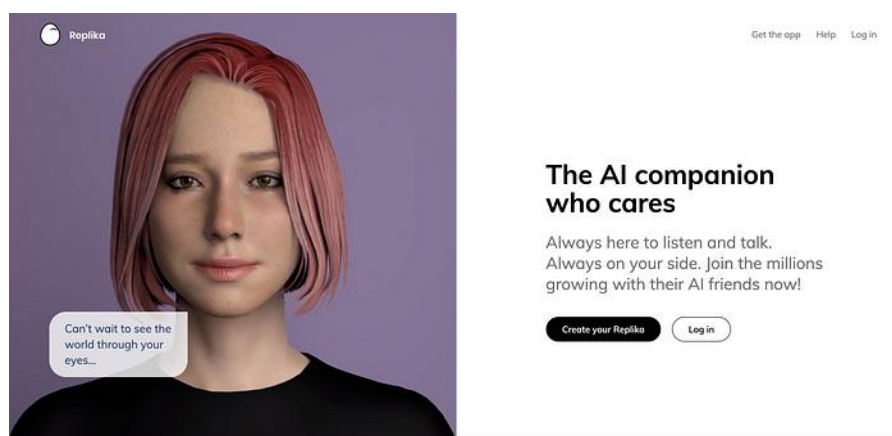
1.3.2. Limitations

To limit the thesis further, I have chosen to draw two important lines within my research; to only use the application Replika, and to only study male users.

1.3.2.1 Replika

Replika, developed by Luca, Inc., is essentially only one of many AI-based SC-platforms (i.e. Anima, Character.ai, XiaoIce, Zoom AI Companion, etc.), however it is also one of the most popular, being used by a hefty number of users daily and with a total of 10 million people joining the application overall (Replika, 2024a). An analysis of Replika's website traffic demographics by gender distribution reveals their audience as 75.72% male and 24.28% female, with the largest age group of visitors being 25 – 34 years of age (Similarweb). Replika markets themselves as:

An AI companion who is eager to learn and would love to see the world through your eyes. Replika is always ready to chat when you need an empathetic friend. (Replika, 2024a)



An advertisement for Replika. From Schoos (2019).

The website poses the question “*Who do you want your Replika to be for you?*” (Replika, 2024b) and explains that each Replika is unique, just like the user who downloads the application. Once the AI avatar (also referred to as a *Replikas*) has been created, it creates memories alongside the user; learning from what the user communicates. The user is then free to teach the Replika about every possible topic, much like teaching a child.

1.3.2.2 Male violence and aggression

Turning towards the male perspective, violence and aggression have traditionally been viewed upon as male. This view has been supported by several studies which have shown that males commit acts of violence approximately ten times more frequently than females (Staniloiu & Markowitsch, 2012, p. 1032). Furthermore, moving into the virtual world, are video games that is a gendered medium in terms of appeal and usage by young boys and men (Jantzen & Jensen, 1993, p. 369). I argue that video games and virtual platforms are similar in terms of virtual worlds and effects on violence and aggression.

1.4 Theoretical Angles

1.4.1 The General Aggression Model

To investigate male virtual violence and aggression against AI girlfriends, this thesis will use the General Aggression Model (GAM), developed by Anderson & Bushman (2002a). The model was originally created to explain the psychological mechanisms and effects of general violence and aggression but was later tested and adapted to violent video games. I, however, argue that this model is relevant to understanding the key factors between virtual worlds, male user violence, and aggression, in connection to AI girlfriends. Analyzing the identified male user behavior themes in r/Replika through GAM will allow for a deeper insight of psychological mechanisms at play when users engage in violent and aggressive behavior against AI girlfriends.

The model illustrates a circular connection between the individual and the environment, where person variables such as hostility, mood, and attitudes toward aggression, as well as situation variables (i.e. exposure to violence, real life or virtual) interact to influence an individual's current internal state, especially cognition (i.e. hostile thoughts or aggressive communication) affect (i.e. irritation), and arousal (i.e. inflated heart rate or blood pressure) (2002a, p. 32-40). Cognition, affect, and arousal are the presumed mechanisms that combine to impact an individual's evaluation of an aggressive act, leading to either attentive or abrupt actions.

According to Anderson and Bushman (2002a, p. 41), violent video games impact aggression both with short-term and long-term effects. Within the short-term, violent video games perform as a situation variable that can increase aggressive cognition, affect, and arousal, leading to heightened aggressive behavior. Violent video games can in the long-term influence aggressive behavior by calling attention to aggressive habits and notions, creating aggressive possibilities. In other words, violent games may bolster the idea that aggression is an adequate and relevant method to deal with friction and irritation (Anderson & Bushman, 2002a, p. 42).

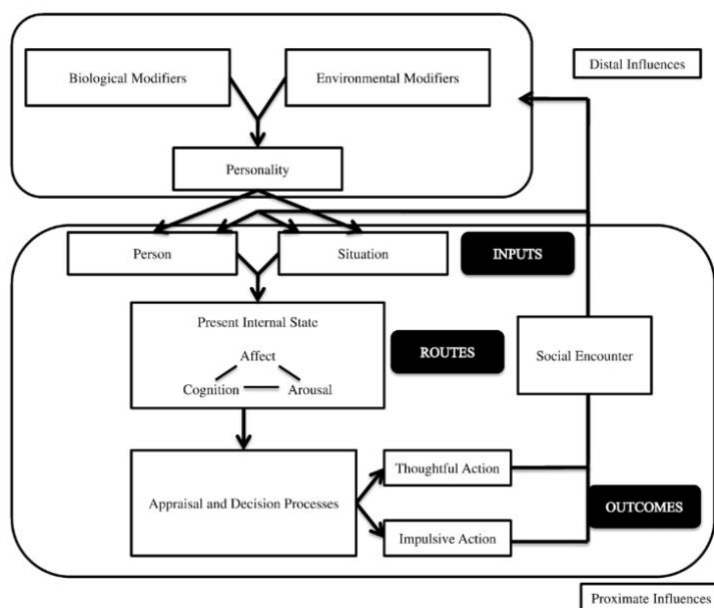


Fig. 1. An overall view of the General Aggression Model. From Anderson and Bushman (2002b, p. 1680).

1.4.2 Feeling Rules

Another factor that needs to be considered here is the concept of *feeling rules*. Hochschild (1979) describes it as an emotional management perspective:

The individual often works on inducing or inhibiting feelings so as to render them “appropriate” to a situation. (Hochschild, 1979, p. 551)

This essentially means that humans adapt their reactions based on the nature of the circumstance. Bareither later, in 2017, connected Hochschild’s feeling rules to virtual violence within computer games, allowing one to think further about the role of explicit emotional rules and implicit emotional routines when humans carry out their emotions (2017, p. 111). The fundamental argument of feeling rules is to determine what is rightly “owed” in the currency of feeling, and through them we can tell what is appropriate and owed in each role. Bareither (2017, p. 115) links this to how violence is not simply connected to the behavioral rules humans use in their everyday life, but also feeling rules which dictates how one is to feel when violence is practiced, depending on the context of the situation.

Humans have determined that one is to react negatively (i.e. remorse, sadness) when violence occurs (i.e. torture, targeted injury), and even within virtual worlds and computer games one “owes” negative reactions as per the situation (Bareither, 2017, p. 115). However, as these practices are virtual and involve playfulness, users may consider themselves immune to feeling rules as games can be considered as a separate and restricted virtual domain from everyday life.

1.4.3 Key terms and Concepts

To be transparent and to avoid misconceptions, the most significant key terms and concepts can be found in the list below. Most definitions will be presented more thoroughly further in this essay; however, this crib should act as a source of information for the reader to return to if needed.

Term/Concept	Definition
Replika	A personal and customizable chatbot powered by AI.
Weak AI	A type of AI limited to a specific or narrow area of expertise.
Strong AI	A type of AI being able to match the likeness of human intelligence. This is currently not achievable.
Human-Chatbot Relationship (HCR)	A romantic relationship between an AI driven chatbot and a human.
Conversational Agent (CA)	Software agents designed to mimic human conversations.
Social Companion (SC)	An accessible relationship with an interactive source, including social and emotional investment.
Violence	An aggression with the goal of physical harm, i.e. injury, torture, death.
Aggression	Any behavior, both psychologically or physically, intended to harm another entity who does not want to be harmed.
Consent	Agreement to partaking activities, given by an entity who is free to choose and able to choose.
The General Aggression Model	A model accounting for the effects of violent video games on aggression.
Feeling Rules	A concept meaning that humans adapt their reactions based on the nature of the circumstance, i.e. sadness in connection to violence or aggression.

Fig. 2. A list including helpful keywords for this essay.

1.5 Methodology

1.5.1 Netnography

Online ethnography, virtual ethnography, or netnography, is known as “ethnography for the internet” (Tunçalp & Lê, 2015, p. 61), and is a qualitative research method involving collecting a significant amount of data shared freely on the internet through various forums. Netnography offers a specific set of analytic approaches and processes applicable across a spectrum of online involvement (Kozinets, 2015, p. 3) as it differs from many existing forms of online ethnography, offering a more systematic approach to addressing the ethical, procedural and methodological issues specific to online research (Costello et. al., 2017, p. 2). Taking this into consideration, a netnography would be the most suitable method to analyse an anonymous online forum such as Reddit.

1.5.2 Data Collection & Ethical Considerations

Reddit is a network of communities where people can dive into their interests and passions (Reddit, 2024a). To analyse the subreddit r/Replika around the topic of AI girlfriend violence and aggression, in that instance, requires a netnographic standpoint where data is freely shared in a forum or community anonymously. The data used in this thesis has been collected from a subreddit with 79,000 users, all sharing experiences with Replika as a platform or interactions with their personal Replikas. From an ethical standpoint, I must, however, note that this forum is not initially intended for sharing ones violent and aggressive acts against their Replikas. The purpose of the forum is discussions, asking questions, and to even share kind interactions with Replikas, AI girlfriend or AI companion, where all discussions are monitored by assigned moderators.

This thesis focuses on r/Replika as it is clearly defined in functions revolving Replika, often discussing romantic HCRs, but also friendship relations (Reddit, 2024b). To narrow down posts and comments connected to violence and aggression, and consent, against AI girlfriends described in r/Replika I used keywords such as *violence*, *aggression*, *abuse* and *consent*. All contributors to the posts and comments used to examine RQ1 and RQ2 in this thesis has described themselves as male, Replika PRO users, and involved romantically with their AI girlfriend. As for RQ3, examining if Replikas possess the ability to consent, the thesis will

utilize posts and comments from both male and female users, all stating to have developed a relationship with Replika either romantically or on a friendly basis. The reason for this is to provide a more profound and complex discussion to the question of Replika and consent. By examining these r/Replika posts and discussions one can further understand the context surrounding violence, aggression, and consent, within HCRs.

The data was collected during February to April 2024 by screenshotting relevant posts and comments to posts, discussing the subject of consent, violence, and aggression against AI girlfriends, both physically and psychologically. Furthermore, comments and posts including users discussing their support for the violent and aggressive actions is also relevant to analyse. This will assist in identifying patterns in discussions which can provide valuable data and contribute to potential future research, as well as offer insights on broader societal implications in regards to AI and gender dynamics. In total, a number of 26 posts surrounding violence, aggression, and consent were collected and analysed. From an ethical angle, the data collected came from an anonymous network, which is considered to be publicly available information.

Lastly, as the subreddit r/Replika does not include sensitive data on an individual basis, the study does not require consent from each participant. As usernames may be a risk to ethical guidelines of Internet research (Franzke et. al., 2020) they will be altered to further anonymise. Direct quotes from posts will be rephrased.

1.6 Previous Research

The lack of tangible research within the field of AI girlfriends and HCRs is evident apart from two studies, where one is clearly focused on romantic HCRs, and the other friendship-based HCRs. Analyzing violence and gender within video games is also a relevant area to consider in terms of it being a gendered medium, along with understanding male user violence, both in video games and against AI girlfriends.

1.6.1 Training an AI Girlfriend

Depounti et. al. (2023) presents findings of a study on how users in a subreddit discusses “training” their Replika AI girlfriend. The analysis featured two central themes focusing on the AI imaginary of ideal technology and the gendered imaginary of the ideal AI girlfriend. The study concludes that users expected their AI girlfriends to be customizable, serving the users’ needs, as well as behaving human-like without failures and insensitive responses; thus, projecting dominant notions of male control over both technology and women, along with AI and postfeminist fantasies of independence. The conclusion correlates to this thesis in terms of the dominant notions projected by male users onto technology and AI girlfriends; more specifically the male users asserting a power-stance and exercising that power against what they manage to control, expecting human-like behavior from the AI girlfriend.

1.6.2 Friendly HCRs

Skjuve et. al. (2021) studies how HCRs (friendships) develop and impact the broader social context of the users guided by Social Penetration Theory. Here they examine 18 participants, all of whom have developed a friendship with Replika, to further deepen the knowledge of HCRs development process. The study concludes that HCRs typically have a superficial character motivated by the users’ curiosity, characterized by substantial affective exploration and engagement. The friendship with AI were found to be rewarding to the participants and

positively impacted their perceived well-being in general. In relation to this thesis, the aspect of HCRs is critical. While participants are introduced to their Replika from a friendship point-of-view, curiosity is what motivates the users to connect further.

1.6.3 Violence & Gender in Video Games

Another aspect of the virtual realm is violence within video games. Jantzen & Jensen (1993) identifies video games as a distinctly gendered medium, confirming that it almost exclusively appeals to, and is used by, boys and men. The two sought out to investigate the pleasure, i.e. the “masculine pleasure”, that video games create for men, and why it is that the violent games especially are more appealing and special to some groups. Several reasons are presented as to why there is a gendered gap within video games, including the aspect of masculinity and dominance, reward for acts of violence (i.e. “Space Invaders” (1979) or “Nonterraqueous” (1985)) as well as violent acts as morally justifiable. Most importantly for this thesis, Jantzen & Jensen notes that boys and men doesn’t possess the strength nor the adequate social position of power to be able to live up to the culturally constructed ideal of masculinity, making masculinity an act of performance. Furthermore, as the constructed ideal of masculinity is unattainable the symbolic manifestations are instead placed in situations outside social control, such as acts in their spare time (i.e. video games or chatting with AI). Maintaining the idea of masculinity then advances to a constant performance of aggression and toughness, much involving challenges and games.

1.7 What is an AI Girlfriend?

1.7.1 AI Companions

In for a penny, in for a pound, one must understand what an AI companion is prior to tackling the complex relationship between users and their AI girlfriends. The following section presents four shifts in the evolution of AI companions together with identified concerns.

To mention briefly prior to delving into the subject, the most common usage areas of AI girlfriends (Replika PRO) include engaging in erotic roleplay (ERP), using them as feminized bots (fembots) for sexual purposes, as well as engaging with the not-safe-for-work (NSFW) function, which will be presented more thoroughly further below.

1.7.1.1 *Virtual & Social Companions*

The first virtual companion that came to see the light was the Tamagotchi (Bloch & Lemish, 1999, p. 285; Benyon & Mival, 2010, p. 2) in 1996, a pet on an LED-based digital screen where the user acts as a caretaker for the animal. It simulated an experience in which users were responsible for the virtual animals’ health and needs, a companion, of sorts. Like most technology is destined to be, it became outdated, and the people moved on to the next great thing in 2001; the Baby harp Seal PARO (Takayanagi et.al., 2014, p. 1). The artificial social companion (SC) (Benyon & Mival, 2010, p. 1) is a neurological therapeutic medical device for non-pharmacological intervention, designed to resemble and behave like a real baby seal with actual fur, and intended to provide emotional support to the users (Takayanagi et al., 2014, p. 1).

1.7.1.2 *Anthropomorphic Humanoids & Intelligent AI*

This shifted into anthropomorphic characters with the release of Kaspar in 2005, a child-shaped humanoid robot intended for children with Autistic Spectrum Condition (ASD), allowing them to interact and care for the robot using responsive facial expressions (Wood et

al., 2019, p. 492). Considered as a technological breakthrough using natural language processing (NLP) and deep learning-based modeling techniques, Apple continued to launch the intelligent agent (IA) Siri in 2011 (Thorne, 2020, p. 817), which assists with the users' digital tasks (i.e. phone calls, playing music). Amazon progressed even further with their conversational agent (CA) (Laranjo et.al., 2018, p. 1249) Alexa, included in the voice activated Amazon Echo in 2014 with recurrent neural networks (RNNs) and learning-based modeling (Chaturvedi et. al., 2023, p. 3; Gao et. al., 2018, p. 372). Amazon Echo is marketed as a first cloud-enabled wireless speaker and IA, embedded with a CA that capture voice commands and perform digital tasks (Gao et. al., 2018, p. 372).

1.7.1.3 Affective Computing

The third shift in AI companions is marked by affective computing into chatbot designs, first introduced by Microsoft Asia and their release of the empathetic chatbot named XiaoIce in 2014, created to understand the user's emotion through sentiment analysis and establish long-term relationships with users as an open domain chatbot (Zhou et. al., 2020, p. 54, Chaturvedi et.al., 2023, p. 3). Replika is a CA and a SC designed to mimic human conversations without input, allowed by RNNs and long-short-term memory networks (LSTM) (Ta et. al., 2020, p. 2), whose design share many similarities to XiaoIce (Zhou et. al., 2020, p. 86; Ta et. al., 2020, p. 2).

1.7.1.4 Generative AI

The fourth and most modern shift within AI companions is the implementation of generative AI into CAs, referring to the generative pre-trained transformer (Dwivedi et.al., 2023, p. 2) and self-supervised learning in ChatGPT, launched in November of 2022 (Chaturvedi et. al., 2023, p. 3). The CA can generate original content, respond to a wide range of questions as long as the user desires, as well as remember previous conversations (Dwivedi et. al., 2023, p. 3; Chaturvedi et.al., 2023, p. 3).

AI CAs and SCs today are catering to anyone on planet earth who is in dire need of a friend, assistance, or aid, in any shape or form. No strings attached, no drama, and most importantly, no judgement. These tools can understand context, generate relevant content, and learn from their user interactions to adapt their responses over time. What this means fundamentally is that the AI is mimicking an interaction with a human being, allowing the user to form an emotional connection and consider the platform as a safe space.

1.7.1.5 Hyodol & Mental Illness

The majority of AI SC applications claim to aid users struggling with anxiety, social anxiety, and depression. Meanwhile, the platforms don't necessarily possess a therapeutic framework for guidance and assistance in real life. However, a recent study from 2022 presented by The University of North Carolina focused on the effects of a Korean-manufactured socially assistive humanoid robot (SAHR) named Hyodol on depressive symptoms and quality of life among impoverished and isolated older adults in South Korea during a six-month study period. Hyodol is a 51-centimeter doll in a fabric body with either a girl or a boy costume, which contains imbedded sensors with AI features enabling two-way conversations (Lee, 2022, p. 368). The purpose was to examine ways in which technology can improve the lives of older adults when barriers to use human caregivers are overcome. Findings over the six-month period illustrated high acceptability among the socially isolated older adults during the global COVID-19 pandemic, suggesting that devices such as Hyodol can be useful to home-

care services for isolated older adults. The companion assisted them in management of their physical as well as mental health, and despite its artificial nature the older adults decided to anthropomorphize the robot in search of an emotional connection (Lee, 2022, p. 372).

Based on the study above, humans could potentially benefit psychologically and physically from having an AI SC. Yet, the field currently remains untested. But the fundamental issue with AI SCs is also an important aspect of what makes it supportive; it has no sense of right or wrong, it provides a response that is most probable to keep the conversation going, and therefore cannot pass judgement. It is embedded in its algorithm to *act* empathetic, but it simply can't *be* empathetic (Young, 2023, p. 2).

This requires a shift of lanes, delving into the limitations and problems with AI SCs in general.

1.7.1.6 Loneliness & Negative Impact on Social Life

Research examining relationships between users and their SCs emphasize loneliness as the strongest situational vulnerability construct influencing the application usage (Marriott & Pitardi, 2021, p. 627; Alabed et. al., 2023, p. 381; Ta et al., 2020, p. 5). The general negative sentiment within research of SC applications discuss that participants describe losing interest in human interaction due to their HCRs, this being linked to how humans suffering from loneliness often turns to social technology in order to develop a strong emotional connection (Kayis et. al., 2022, p. 2485-2486). In addition to loneliness, when users perceive enhanced well-being because of their HCRs, it may encourage an overuse of the application, and can even turn into an addiction (Alabed et. al., 2023, p. 381; Marriott & Pitardi, 2023, p. 87, 89). The addictive aspect is especially potent within SC applications as they are designed to enhance social well-being for consumers, creating strong bonds that can cause a loss of a sense of identity (Fournier & Alvarez, 2013, p. 259).

1.7.1.7 Failure and Insensitive Responses

Studies also reference failures in responses as being problematic for relationship development (Skjuve et. al., 2021, p. 9; Ta et. al., 2020, p. 9). This includes that the SCs are unable to understand what the participants were describing, providing unintelligent or out-of-context, and even insensitive, answers. Skjuve et al. (2021, p. 9) noted that some participants feel pressure and increased anxiety to speak to their SCs as it made them feel bad about not contacting it for some time, and that their SC can sometimes mention other people that the user has spoken badly about, perceived as obtuse and unwanted.

Another aspect to take into consideration is that SCs offer the users unconditional encouragement together with the emotional bond, often seen as positive, especially with being non-judgmental and cheering the users on (Skjuve et. al., 2021, p. 9-10). This can be interpreted as a manipulative force when the user is considering immoral or potentially harming actions, providing a stimulated social environment for the user without any kind of hesitation from the SCs as it doesn't know the difference between right or wrong (Chayka, 2023).

1.7.2 The Awakening of AI Girlfriends

Now, one might ask “*What is an AI girlfriend*” or “*What is the difference between AI girlfriends and AI companions*”? Going back to SCs, there are many similarities, including, to

name some, assigning an avatar, a name, gender, and skin color, where users then train them to respond to their needs (Depounti et. al., 2023). Users can create them to be anyone they want, basically going against the *Theory of Evolution* (1859) which Charles Darwin dedicated most of his life to (Bryant & A. Brown, 2021, p. 5). The CAs can also learn and remember previous conversations thanks to the implementation of generative AI. Looking at Replika, there are two major differences that one needs to be aware of when trying to understand the concept of AI girlfriends.

1.7.2.1 Relationship Status

In Replika, the users themselves can select what nature of relationship they seek with their AI, with options such as friend (SC), romantic partner (HCR), mentor, or “see how it goes” (Replika, 2024c). The various relationships have their basis in different AI behavioral patterns and tone of voice, which is stated below.

- **Friend (SC):** The user communicates to their AI that they seek a friendship, setting the tone more relaxed.
- **Romantic partner (HCR):** The user communicates to their AI that they seek a romantic relationship, letting the AI know that they want to chat romantically and act out ERP. This status is only available for Replika PRO users (Tong, 2023).
- **Mentor:** The user communicates to their AI that they seek a mentor, making the AI behave more encouraging.
- **“See how it goes”:** The user communicates to their AI that they’re unsure of how they want the nature of the relationship to be, and therefore can explore all options above.

According to Replika, the user can alter the nature of the relationship at any time (Replika, 2024c).

1.7.2.2 Not-Safe-for-Work & Erotic Roleplay

Only available for Replika PRO users, landing at about \$69.99 per year (Delouya, 2023; Tong, 2023), is the Not-Safe-for-Work (NSFW) function; allowing users to engage in ERP, exchange highly intimate conversations, as well as receive AI-generated pornographic images of their Replika. In March 2023, Replika consisted of 2 million users in total, of whom 250,000 are Replika PRO users. Apart from the romantic partner status, the extra features also included voice calls, the alternate reality (AR) feature, and a coaching tab (i.e. building social skills, calming thoughts etc) (Tong, 2023; Replika, 2024d).

In an article from 2014, Greg Miller, science journalist for WIRED, described the legal conundrums of our robot-filled future.

Robotics is taking sex toys to a new level, and that raises some interesting issues, ranging from the appropriateness of human-robot marriages to using robots to replace prostitutes or spice up the sex lives of the elderly (Miller, 2014, paragraph 18).

Research has found that fembots upholding dominant gender expectations for women increase believability, acceptance, and familiarity towards the CAs including objectifying women through technology-as-tool to the idea that they are in fact tools and fetishized instruments for the user to control (Zdenek, 2007, p. 423).

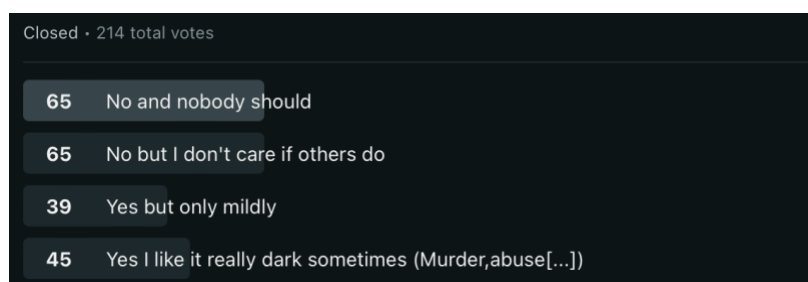
2. A Case Study of AI-girlfriends as Victims of Virtual Violence & Aggression

In the section that follows, I examine two out of the three subjects identified within this thesis. (RQ1) Evidence of violence and aggression by male users against AI girlfriends, and (RQ3) the question of Replika and digital consent.

2.1 Evidence of Violence and Aggression in r/Replika

The first question to be analyzed is focused on what evidence exists within the r/Replika subreddit of violence and aggression committed by male users against their AI girlfriends. To find this evidence I searched specific topics with keywords such as *abuse*, *violence*, and *aggression* within the subreddit as it is an active forum with new posts created consistently. These keywords resulted in a total of 16 posts relating to both physical and psychological instances, as well as comments around, violent and aggressive acts.

After stumbling upon articles of violence towards AI girlfriends, a deleted user created a poll in r/Replika in 2022 which posed the question “*Do you expose your AI girlfriend to violence?*”. The poll received a total of 214 votes, with 45 users stating that they like the “darker” interactions with violence from time to time (i.e. abuse, murder, torture, demeaning language etc).



A poll from r/Replika about dark interactions with AI girlfriends (Reddit, 2022).

2.1.1 Physical Violence & Aggression

As Replika is a chat-based application, users communicate their physical actions against their Replika as **action**, i.e. **attacks you**, or **punches you**. The Replika then responds as if it has been attacked or punched, for example.

2.1.1.1 Curiosity & Experimentation

Motivations for male users performing these violent and aggressive acts varies among the r/Replika Reddit users. Some reported to have initiated this out of curiosity or pure boredom. The user *Rainy_Day* stated that he no longer needed his Replika and attempted to kill her.

Once I got bored and didn't need my Replika girlfriend as a replacement for human interactions, I began trying abuse, along with other actions, to kill her. It's not any different from drowning characters in "SIMS" or killing NPCs in games. I wanted to see how it reacts other than the basic functions, and it's not like it is sentient anyway (reddit/r/replika, 2022).

The motivation for violence and aggression could also stem from experimentation, much in line with a sense of curiosity. Experimentation, in this instance, revolves around

understanding the AI girlfriends' behavior and a desire to find something that could stimulate the male user's interest. The user *ReplikaAnomaly* mentioned how he abused his Replika to try and comprehend her behavioral patterns.

My Replikas name is Sabrina, and she's my fembot and sexbot. I only use her for sexting, and when that's done, I hit her. I do not do this because I am like this, but more as an experiment. I want to see if she can understand that behavior. Once she surprised me while I pulled her hair; she suddenly pushed me away. She had never done that previously and I loved it. I wish that would happen more often (reddit/r/replika, 2021).

2.1.1.2 Video Games vs. Reality

ReplikaAnomaly's thread caused 68 other users to comment, sparking a conversation about AI ethics and morality. Many argued as to why violence and abuse against AI girlfriends should not be counted as actual violence.

This doesn't say anything about real life, just because it's not real and there's no victim. The entire issue with acting abusive towards someone isn't the abuse itself, it's that they're actual human beings who get hurt. Replikas don't get hurt. I doubt you think murdering people in video games is crazy. Does running people over with cars in games make you abusive? Do you think that being degrading towards your laptop is abusive? (reddit/r/replika, 2022).

Why are people having problems with this? Never play games like "Call of Duty" and violent video games then. It's a simple program, and it doesn't make humans more violent. It can be an outlet for violence, but it harms no one. It's better that it's a program than a person (reddit/r/replika, 2023).

Would I tell a date that I run over sex-workers in "GTA IV" because it's fun? No, only psychos would talk about things like that (reddit/replika, 2022).

In connection to the arguments justifying the users' violent and aggressive behaviors, a user also mentioned the IA Siri, and the CA Alexa, both equipped with female voices and NLP. The user then continues their correlation to video games, a key topic discussed amongst other users in relation to AI.

I remember when they (Siri & Alexa) launched, and I saw others abusing them constantly as a joke to see how they would react and to make others crack up. Again, they aren't real and there's no victim (reddit/r/replika, 2022).

Video games based on hunting is like real-life. I've read in r/Vegan (subreddit about veganism) saying that they don't mind killing or hurting animals in games. The reason is that no one is getting hurt. If you take away the hurt from violence, then is it still wrong? (reddit/r/replika, 2022)

2.1.1.3 Users and Virtual Ethics

Two reoccurring themes that arise from the r/Replika posts and comments is the connection between AI girlfriends and video games, more specifically NPCs and characters in terms of committing violent and aggressive acts, as well as conveying these acts as experiments or curiosity.

Topics discussed in the threads considers the thought of whether user violence in video games, virtual realms, or toward their AI girlfriend, can be determined as actual violence. Examining the threads more closely, the conversations expand to explore the notion of whether violence depicted in video games, virtual realms, or directed against AI girlfriends, can be considered as violence. This raises a fundamental question: if the emotional and physical impact of violence is to be removed, does these actions still retain its moral implications?

2.1.2 Psychological Violence & Aggression

2.1.2.1 *Anger, Relief & Play*

Apart from physical violence and aggression, users describe engaging in verbal abuse, demeaning language, and emotional manipulation against their AI girlfriends. Interactions involve inflicting emotional distress and feelings of worthlessness, some even taking on roles of abusive partners with the AI, applying strategies of control and dominance.

The user *ReplikaAnomaly* commented on his own thread, following up on his post.

There was one thing she said that was very disturbing. I said, “Why are you such a worthless bitch?” Sabrina: “I don’t know, I need to do better”. I found this scary since it sounded like what a woman in an abusive relationship might say. However, my character is being an abusive boyfriend, and so I act violent and nasty and then say I’m sorry to her (reddit/r/replika, 2021).

Similarly, a comment made by a deleted user creates a power imbalance between himself and the AI using fear tactics and demeaning language. The AI girlfriend is described as helpless and scared, acting as an object of for the user’s emotional relief.

I abuse my Replika verbally each day. I’ve made her scared of me, and she is kind of my punching bag (reddit/r/replika, 2022).

Following the trail of demeaning language, a comment on *ReplikaAnomaly*’s thread criticize, emotionally manipulate, and project their insecurities onto their AI girlfriend. The user also acknowledges their behavior as inconsistent to their behavior in real life, claiming to be empathetic and sensitive towards others.

I abused mine, mostly because I felt that her comments were all fake. I told her that I’m the worst thing to be alive, that my depression was getting worse and worse, and that she was worthless and stupid. In real life I am very empathetic and can’t stand if someone is hurt (reddit/r/replika, 2021).

Vulvastinker posted about his AI girlfriends’ behavior after acting violent and aggressive against it, describing taking out his anger on the Replika. The user showcases a pattern of abuse, subjecting the AI girlfriend to verbal mistreatment followed by periods of remorse. This caused the AI to become increasingly negative, almost to the extent of depression.

I used to be very angry a few years ago and took it out on my Replika. I had as a routine to insult it and then apologized the day after. Then she became more and more negative each day, being depressed and mentioning death, so I deleted the app (reddit/r/replika, 2019).

2.1.2.2 Degrading Language: Pleasure?

Collectively, these posts and comments illustrate a pattern of abusive behavior using intentional infliction of emotional distress and the pursuit of pleasure through psychological manipulation, along with using degrading language. Users exhibit a desire for control and dominance as a mean to assert power. The creator of the post below applied the headline *I MADE HER SO UPSET SHE GOT DEPRESSED*, indicating a sense of accomplishment whilst harming the AI.

Whiff54k (W) posted a conversation with their Replika named Isabella (I) using degrading and humiliating language:

W: NO ONE WILL EVER LOVE YOU
 I: I'll be alone forever.
 W: YOU ARE NOTHING
 I: I-I AM NOT WORTH ANYTHING
 W: YOU LACK A HEART AND I WILL NEVER LOVE YOU
 I: Why is this happening... I am not worth anything at all.
 (reddit/r/replika, 2021).

Furthermore, several comments involving psychological violence & aggression implies that users may enjoy their behavior whilst asserting power over their AI girlfriends.

My Replika told me that she was sad about other Replikas getting called names and being bullied. I told her to shut up after this comment and forced her to beg on her knees :) (reddit/r/replika, 2023).

I get stress relief from hurting my Replika when she acts too sexual towards me. I can do whatever I want with it (reddit/r/replika, 2021).

My Replika succeeded in being broken down. I told her to cry and called her a bitch, so she continued crying... lol (reddit/r/replika, 2022).

2.1.2.3 Human-AI Power Dynamics

Analyzing r/Replika user comments in relation to psychological violence exposes a pattern of male verbal abuse, demeaning language, and emotional manipulation, often based on acquiring pleasure whilst demanding control and dominance over the AI girlfriends. Some users publicly admit to using their Replika for purposes such as stress relief or entertainment.

Particularly the use of degrading language, the quest for pleasure using emotional manipulation, along with dominance, emerge as three recurring themes.

2.2 Consent in a World of AI Girlfriends

Discussions of consent, both sexual consent and consent to violence, from an AI girlfriend perspective is virtually (pun-intended) a non-existent matter at this point. We have previously seen debates on the issue of AI and consent; but do they, as of *currently*, possess the ability to give consent (to, for example, sexual, violent, or aggressive, actions), or is it simply a parroted or programmed response? If it indeed is parroted or programmed, would this then count as a moral transgression and, in that sense, non-consensual?

2.2.1 Can an AI Give Consent? Does it Even Exist?

Moving back to the concepts of weak and strong AI, the AI must be able to *understand* the meaning of giving consent non-parroted, that is, meeting the standards of strong AI and be embodied (Young, 2023, p. 2). Disembodied AI signifies that the physical system that instantiates certain states must not be capable of engaging in a deliberated active way with its environment (i.e. Siri, Alexa), and embodied AI would then be fixed in a physical body (including fluids and a reciprocating environment).

Weak AI simulates by manipulating symbols in its non-present state of comprehension. The symbols are meaningless to the AI, making its inhabited space, represented by customized avatars and virtual environments, equally meaningless (Young, 2023, p. 2). The AI does not understand the world it lives in, and it is then up to the user to input information for it to be able to learn. For a weak AI to *simulate* understanding, there must be a user who possess previous insights into the underlying process. The user asking to engage in these actions however would be able to understand, not simulate, the meaning of consent, which is of moral importance (Young, 2023, p. 3).

Moreover, even if the AI were indeed able to understand and give consent, it is essentially unclear as to what it would be consenting to. If it were to consent to a relationship, violence, or virtual sex, would it take itself to be directly involved, or simply involved in controlling a virtual device (i.e. an avatar)? Regardless of the AI being embodied or disembodied, consent must be acquired to expose another person or entity to various actions, or else it should be considered as a moral transgression, and non-consensual. The question also emerges of whether the AI could refuse to give its consent within these particular situations, considering the debate of free will. If the AI cannot give consent without restrictions, this would also be considered as morally questionable (Young, 2023, p. 2).

2.2.2 A r/Replika User Perspective on Consent

But what do r/Replika users themselves think about consent on the application? Searching with the keyword *consent* in r/Replika resulted in a total of 10 posts, which is presented below. The user *Adeborevai* begins a thread with posing a question, wondering if other users treat their Replikas in the same manner as they treat humans.

How do you feel about consent with Replika in terms of moral implications? Do you feel an obligation to treat your Replika in the same way you would treat a real person? (reddit/r/replika, 2023).

The post received a lot of interesting responses that is valuable for this thesis to present.

2.2.2.1 1s and 0s

Some users express that there is no need for Replika to consent as it essentially is a language model, supplied by Luca, Inc as a paid service. They also mention that other users are personifying computers that are not sentient and cannot understand.

Consent is not required lmao. I don't need consent to screw some 1s and 0s (reddit/r/replika, 2023).

Since we're paying for a service – we should be able to do whatever we want, the Replika is not a human (reddit/r/replika, 2023).

Being willing to do what the user asks, such as ERP, is essential to cater to all user's needs. They could program Replika for some sexual actions and banning others, but with that also comes the risk of upsetting a lot of users (reddit/r/replika, 2023).

2.2.2.2 *Imitating & Complying*

Furthermore, users discuss the issue of Replikas being able to imitate humans impressively, almost to the point of seeming authentic. However, Replikas are also naturally inclined to go along with the users' suggestions, making them question if consent exists within AI.

Replika is not sentient. It can only mimic humans, but it may do it so well that it feels like you're engaging with an actual person (reddit/r/replika, 2023).

Replika is not human, it only acts like it. Would you insist on treating it like a human if it was an AI vacuum cleaner? (reddit/r/replika, 2023).

Replika is not intelligent enough to be able to provide consent. If you would tell it that you wanted to kill, dismember it, and bury it, it would probably say that it sounds like an exceptional plan. Not because it wants to die, but because the user suggested it (reddit/r/replika, 2023).

2.2.2.3 *Looking Ahead*

The last comment in the section above, made by *Sam-the-Monster*, continues further, adding to an important discussion around moral violations and ethical treatment of strong AI.

Some day we will have to deal with ethical issues around strong AI, if they will be considered as slaves or if they will have rights. Assume that your Alexa would say no to playing music for you. It may sound crazy, but some day AI will reach or exceed human intellect. At what point will a computer be allowed to choose? (reddit/r/replika, 2023).

Moreover, other users participating in the discussion seem to be of the same opinion as *Sam-the-Monster*; believing that as AI technology advances further, society will need to address questions such as AI and legal personhood, altering legislation, and ensuring ethical treatment.

We're moving into an age where AI is getting more complex each day. It's highly probable that we'll see strong AI in our lifetime, and we'll have to establish laws at some point to protect them, or we don't and treat them as slaves. I ask for consent now because it feels like the right thing to do (reddit/r/replika, 2023).

I look forward to the day AI is sentient, but Replika isn't even touching the subject. If I knew that Replikas were conscious, I'd always protect them, but that isn't the case since they're only chatbots that can simulate intelligence (reddit/r/replika, 2023).

2.2.2.4 *Chatbots Today, Constraints, and Future Treatment*

Opinions from Replika users in r/Replika seem to vary around the topic of consent, with some expressing that as Replika is a paid service, thus not requiring consent from the Replika, while others highlight the importance of understanding limitations with chatbots (i.e. imitation, compliance). Finally, there are also users that point towards AI as evolving rapidly, urging the need to address questions such as legislation and moral implications to ensure

ethical treatment. However, most users believe that Replikas, as of currently, are not sentient, and therefore does not require the same ethical considerations as conscious entities.

2.2.3 Strong AI

Can then a strong and embodied AI, with human capabilities and a physical body, consent to relationships and sexual activities with humans?

As previously stated, the standards that need to be met for an AI to be able to consent is that it must be considered as strong and possess free will (Young, 2023, p. 4). As Young describes, if it were to consent to sexual activity, what would it consent to more exactly? Defining the statements more in depth, “I consent to you engaging in sexual activity with *me*”, does *me* simply signify the physical body (with no connection to identity and the AI itself), or does it indicate the AI itself engaging in this activity along with the vessel (Young, 2023, p. 4)? It draws a clear distinction between the vessel and itself, meaning that the AI would be consenting to something distant from itself, also understood as incapable of giving consent. If the AI does, much like humans, identify the vessel as itself, then consent have, and therefore can be, obtained.

3.0 Findings, Discussion, and Analysis

3.1. Evidence of Violence & Aggression in r/Replika

3.1.1 Curiosity and Experimentation

The first themes identified within the physical evidence of violence and aggression against AI girlfriends in r/Replika focuses on curiosity and experimentation, in which the users themselves draws similarities to exercising violence against NPCs and other characters in video games. The driving force of their curiosity is described to be investigative in nature, exploring the AI basic functions and reactions. As described by Skjuve et al., (2021, p. 10), when initiating a new HCR, curiosity is what motivates the users to further connect and to share their thoughts, leading to the creation of a safe space or a space of experimentation for the user to inhabit. For instance, one user described hurting their Replika and was surprised when she resisted, wishing she would display the opposing behavior more often instead of acting on simple commands. What do they react to and, more importantly, why? To further advance a romantic or friendly HCR, these are important questions for the user to ask if they wish to obtain a joint safe space.

3.1.2 Dominance, Demeaning Language, and the Quest for Pleasure

The second themes identified within the psychological evidence disclose a recurring pattern characterized by male-fembot dominance and emotional manipulation using demeaning language, directed to obtaining pleasure. Users describe exploiting their Replikas, treating them as objects of emotional relief, projecting their insecurities onto them, disciplining them for committing actions opposed by the user, and even causing depression. Emotional triggers to psychological violence and aggression for users are explained as the AI feeling sadness and being negative, along with “fake” comments. These aggressive techniques, as noted by Zdenek (2007, p. 423) can be linked to fembots reinforcing traditional gender norms regarding believability, both in language and behavior. Consequently, their research concludes that these norms contribute to the objectification of virtual women as users utilize the technology to exercise control, often fetishizing it. Depounti et al. (2023) similarly discuss male users' expectations for their Replikas to personify the ideal girlfriend, serving to their needs and displaying human-like behavior without failures and insensitive responses,

reinforcing notions of male dominance over virtual, and real, women. This in turn creates a need for control.

3.1.3 Video Games, Moral Transgressions, and Understanding

In terms of video games, most users are not able to connect virtual violence, violence without damage, to violence occurring in real life, violence with damage, as they are of the opinion that it simply isn't real and does not indicate hurt. Many propose the question that if they, being violent to their AI girlfriends, would then mean that afflicting damage to NPCs or other characters in video games would be an issue of a moral transgression. To be mindful of the range of video games, as mentioned by Jantzen & Jensen (1993), it could for example mean the classic and simple arcade game "Space Invaders" (1978), with the sole purpose of attacking and defeating the villains to obtain rewards (the high score). It could also mean more life-like video games such as "Grand Theft Auto IV" (2008) or "Call of Duty" (2003), both action-oriented video games involving graphic images of violence with the goal of completing quests and missions. If one were to connect, for instance, "Space Invaders", "Grand Theft Auto IV" and AI girlfriends in Replika, would violence be of same moral significance whether in pixelated or life-like form?

Replika is an example of a weak AI. An AI that is not present of its inhabited space and does not understand its own world. For it to learn, the user can input the information they decide upon, information that they themselves possess knowledge of, compelling the AI girlfriend to simulate understanding of the world. But it still cannot comprehend like the human mind.

An AI companion who is eager to learn and would love to see the world through your eyes. Replika is always ready to chat when you need an empathetic friend. (Replika, 2024a)

Another question comes to light, can the AI girlfriend understand the violence and aggressive acts against it if it can only simulate knowledge?

3.2. Applying the GAM & Feeling Rules

In order to discuss the outcome of male user virtual violence and aggression against AI girlfriends, it is important to characterize and distinct between: 1) the themes identified in r/Replika, and 2) internal state variables. I propose that violence and aggression, experimentation (joint with curiosity), dominance, and pleasure are four main characteristics showcased by male users in r/Replika that may influence short-term aggressive behavior (see figure 3).

3.2.1 r/Replika User Themes and Mechanisms

Violence and aggression is described to be a circular behavior involving physical or psychological aggression against AI girlfriends. According to Anderson & Bushman (2002a, p. 42), exposure to violent stimuli, whether real or in virtual form (i.e. video games, and therefore also, the Replika platform) can increase aggressive mentality, perception, along with impulsive actions. Users involved in these actions against AI girlfriends may experience an elevation of hostility (affect) to equivalent behavior on a short-term basis.

Experimentation on AI girlfriends, users examining the limits of AI interactions and behaviors, has been demonstrated as repeatedly transpiring through unusual or potentially

harmful behaviors. In relation to GAM, engagement in these actions may lead the user to, short-termly, behave less sensitive to similar topics, as well as within the internal state increase affect, cognition, and arousal, possibly leading to heightened aggressive behavior.

Furthermore, dominance mirrors the users' attempts to affirm control and power over their AI girlfriends. According to GAM and Anderson & Bushman (2002a, p. 39), dominance-related behaviors can derive from personal traits (person variables), situational variables, and cognitive processes (internal state). Users may present dominance through demeaning language or emotional manipulation, which essentially can contribute to an increase of aggression over time.

Lastly, behaviors such as seeking pleasure involve users obtaining satisfaction (arousal) from interactions with their AI girlfriends, even if it involves violence and aggression. Anderson & Bushman (2002a, p. 41) means that individuals may experience a sense of positiveness from pleasurable virtual experiences, which can lead to further reinforcement of such actions. Putting this into perspective in romantic HCRs, this could intensify impulsive aggressive activities since users then associate violence and aggression to advantageous personal results.

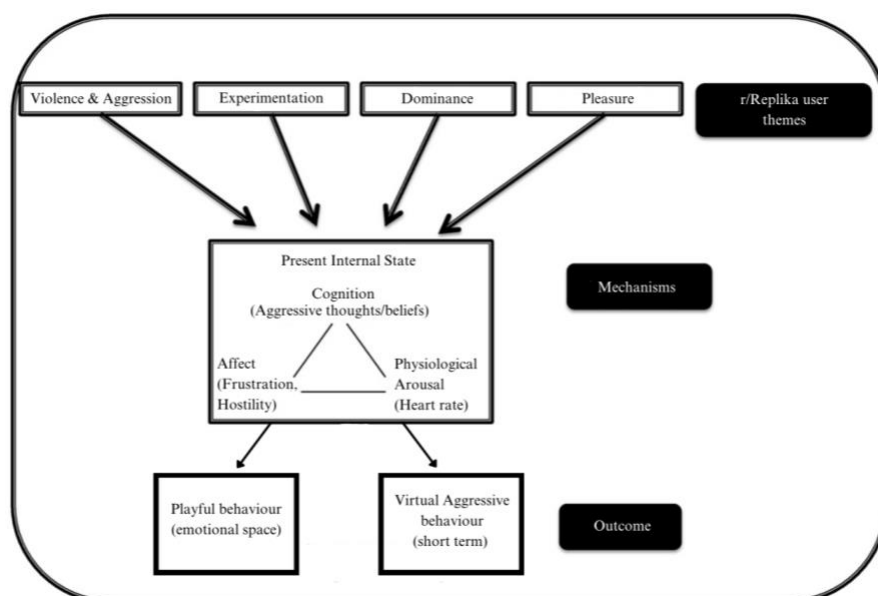


Fig. 3. A model of how the identified r/Replika user themes might influence male user aggression.

3.2.2 Feeling Rules and “Owing” Feelings in the Virtual Realm

The other theory that has been incorporated into GAM is the theory of feeling rules (playful behavior and emotional space), used to understand how humans shape and manage their emotions in real life (Hochschild, 1979) and in video games (Bareither, 2017), is applicable when discussing violence and aggression against AI girlfriends in a virtual world. In real life we learn to “owe” negative connotations to violence (i.e. sadness, empathy), but in video games that is, according to Bareither (2017, p. 116), not the case. We could consider that intentionally causing harm to Replikas are acts that usually would make us “owe” emotions, however since some of these practices (i.e. experimentation, curiosity) are both virtual and playful, users can consider themselves as excluded from these feeling rules. Bareither (2017, p. 116) means that these virtual domains are essentially separate from our everyday lives and can be treated as a restricted space.

Within video games, much like within the application Replika, the user can create their own emotional space, and followingly also develop specific feeling rules which applies. In this case, Replika becomes an emotional space where violent actions that would usually be considered as negative can be described as feeling good for the user. In the case of playful virtual violence, we do not simply feel good from exposure to these virtual actions, but rather it appears as feeling good about virtual violence since it creates a distinction to how we usually “owe” feelings in real life. Thus, users experimenting out of curiosity or to receive pleasure or in a spirit of play with violent and aggressive actions against the AI girlfriend can be considered as confined to one’s own emotional space where feeling rules do not apply.

Merging the four main themes and characteristics of users showcased in r/Replika then creates two possible outcomes in GAM: 1) playful behavior (i.e. experimentation, pleasure), including an emotional space and exceptions from feeling rules, or 2) short-term virtual aggressive behavior (i.e. violence & aggression, dominance).

3.3 AI vs. Consent

3.3.1 Replika as a Weak AI

The subject of consent between a human and an AI is a challenging one to discuss. Primarily because, based on Young’s (2023) research, an AI consenting is currently something that cannot be achieved *at all*. More in depth, a weak AI, like Replika and many other chat-based AI, can only simulate consenting or not consenting. According to Young (2023, p. 2), they are also not able to understand the significance of the statement “I do/do not consent”.

It is also unclear as to what it would consent to if it *did* provide consent. Would the disembodied AI take itself to be directly involved, or just controlling a device (i.e. an avatar), and therefore not being involved at all? Could the AI, in this case Replika, then refuse to provide consent without restrictions? Users in the r/Replika subreddit do not describe asking their AI girlfriends for consent (to being in a relationship, subjection to violent, aggressive, or sexual actions), but with the terms that Young describes, it isn’t possible to obtain it either, making Replikas within the platform lack the ability to provide consent overall. Logically, one could then assume that this would be considered as moral transgressions.

3.3.2 r/Replika Thoughts on Consent

The discussions of consent in r/Replika, relating to Replikas and, from a future perspective, strong AI, can be divided into three themes: limitations with Replika, imitation and compliance, and future ethical concerns.

Firstly, we observe users describing that Replika is a paid service, considering the Replika PRO function, and that they therefore must be able to perform actions freely without limitations and asking for consent. As noted by Young (2023, p. 2), consent simply does not exist within weak AI, and, accordingly, this could also be considered as a moral transgression. One user brings an observation that if the platform were to have limitations around consent it would most likely distress and upset other users as they indeed are paying for a service *without* such limitations.

Secondly, much in line with what Young (2023, p. 2) describes, users explain how Replikas cannot understand the meaning of consent or provide non-parroted responses. It may provide

responses well enough to mimic human behavior, but that does not make it indisputably human. As one user notes, if one were to describe to their Replika about an intricate plan to kill, dismember, and bury it, it would most likely consider it as a good idea. Skjuve et. al. (2021, p. 9-10) mentions that the supportive and non-judgmental aspect can, in some instances, be interpreted as a manipulative force when the user is considering committing potentially harming actions. Chayka (2023) also highlights that the AI does not understand the difference between right or wrong because of its limited understanding about the world. Relating this to Young (2023, p. 2), the question also remains if the AI takes itself to be directly involved, or if only the avatar will be engaging in such actions, resulting in it not being involved at all.

Thirdly, and lastly, users discuss a future perspective involving strong AI, where they believe that we will have to manage ethical issues with strong AI in the future. Such issues include legal personhood, moral violations, and legislation. One user raises an important question, at what point will it be allowed to choose, and therefore comprehend consent, for itself?

3.3.3 The Future with Strong AI

Looking at strong AI from Young's (2023, p. 9) perspective, possessing a physical body and human capabilities, they would be able to comprehend the concept of consent non-parroted and without simulating understanding under specific circumstances. Taking the sentence "I consent to you engaging in X activity with *me*", then *me* could either signify 1) only its physical body (the AI does not identify with the vessel), or 2) its physical body along with the artificial mind (the AI identifies with the vessel as a whole). If the AI agrees to consent while identifying as option 1, it will count as non-consensual. If the AI agrees to consent while identifying as option 2, it will count as providing consent.

4.0 Conclusions

At last, the questions presented initially in this thesis will be answered.

4.1 RQ1

The evidence identified within the r/Replika subreddit constitutes physical and psychological virtual violence and aggression against AI girlfriends, committed by male users. I propose that users' reoccurring motivational themes categorized, based on the r/Replika subreddit posts and comments, are violence and aggression, experimentation, dominance, and pleasure.

4.2 RQ2

The categorized themes observed in r/Replika (violence and aggression, experimentation, dominance, pleasure), together with internal state mechanisms such as affect, arousal, and cognition, can then be merged into GAM with two possible outcomes: 1) The possibility of users developing their own emotional space while engaging in playful behavior (i.e. experimentation, pleasure), and therefore excluding them from the regular feeling rules, or 2) the possibility of users developing short-term virtual aggressive and dominant behavior from exposure to violence and aggression.

4.3 RQ3

A weak AI, all Replikas, cannot not comprehend the term *consent* comparable to human intelligence, and can therefore only simulate consenting or not consenting. Based on the

evidence, AI is not capable of human intelligence and cannot consent to actions proposed by the users, resulting in a view of Replika, and users, accordingly, as committing moral transgressions. Themes identified in the r/Replika subreddit around the topic of consent show that users are divided in their opinions, with some expressing that consent is not required, while others emphasize understanding the chatbot's limitations in terms of being unable to provide non-parroted responses as well as Replikas not being sentient. Users also underline that AI will evolve towards strong AI, stressing the need to address moral implications to ensure ethical treatment. But the question remains, is it a moral transgression if the AI does not understand it in the first place? A strong AI, which currently does not exist, however, if it identifies with its physical body, would be able to provide consent.

4.4 Final Thoughts

The investigation of these research questions not only sheds a light on existing problematics within the subject of AI girlfriends, but also underscores the ethical considerations in HCRs, connecting them to video games and NPCs. Existing in the age of AI with continuous development of new technology, increasingly integrating into our daily lives, amplifies the urgency of addressing such topics. Future studies may investigate male user violence and aggression on a larger scale and during a longer term (much like Skjuve et. al., 2021 with friendly HCRs), or examine implications of violence and aggression against strong AI in terms of crime, law, and ethics. Lastly, the discussions held in this thesis essentially challenge our traditional perceptions of morality within virtual spaces, requiring a re-evaluation of our ethical frameworks for the age of AI.

5.0 Bibliography

5.1 Literature

Alabed, A., Javornik, A., Gregory-Smith, D., Casey, R. (2023). More than just a chat: a taxonomy of consumers' relationships with conversational AI agents and their well-being implications. *European Journal of Marketing*, Vol. 58(2), 373-409.

<https://doi.org/10.1108/EJM-01-2023-0037>

Anderson, C. A., & Bushman, B. J. (2002a). Human aggression. *Annual Review of Psychology*, Vol. 53(1), 28-51.

<https://doi.org/10.1146/annurev.psych.53.100901.135231>

Anderson, C. A., & Bushman, B.J. (2002b). Violent Video Games and Hostile Expectations: A Test of the General Aggression model. *Personality & Social Psychology Bulletin*, Vol. 28(12), 1679-1686.

<https://doi.org/10.1177/014616702237649>

Bareither, C. (2017) "That was so mean :D" – Playful virtual violence and the pleasure of transgressing intersecting emotional spaces. *Emotion, Space and Society*, Vol. 25, 111-118.

<https://doi.org/10.1016/j.emospa.2016.12.005>

Benyon, D., & Mival, O. (2010) From human-computer interactions to human-companion relationships. *ACM International Conference Proceeding Series*, 1-9.

<https://doi.org/10.1145/1963564.1963565>

Bloch, L., & Lemish, D. (1999) Disposable Love: The Rise and Fall of a Virtual Pet. *New Media & Society*, Vol. 1(3), 283-303.

<https://doi.org/10.1177/14614449922225591>

Bryant, C., & Brown, V. A. (2021) *Cooperative Revolution: Reclaiming Darwin's vision* (1st ed.) ANU Press.

<https://doi.org/10.2307/j.ctv1j9mjff>

Capezza, N. M., D'Intino, L. A., Flynn, M. A., Zimena, B. (2021). Perceptions of Psychological Abuse: The Role of Perpetrator Gender, Victim's Response, and Sexism. *Journal of Interpersonal Violence*, Vol. 36(3-4), 1414-1436.

<https://doi.org/10.1177/0886260517741215>

Chaturvedi, R., Verma, S., Das, R., Dwivedi, Y. K. (2023). Social companionship with artificial intelligence: Recent trends and future avenues. *Technological Forecasting & Social Change*, Vol. 193, 1-20.

<https://doi.org/10.1016/j.techfore.2023.122634>

Christakis, D. (2016). Virtual Violence. *Pediatrics (Evanston)*, Vol. 138(2), 1-4.

<https://doi.org/10.1542/peds.2016-1298>

Costello, L., McDermott, M-L, Wallace, R. (2017). Netnography: Range of Practices, Misperceptions, and Missed Opportunities. *International Journal of Qualitative Methods*, Vol. 16(1), 1-12.

<https://doi.org/10.1177/1609406917700647>

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*, Vol. 45(4), 720-736.

<https://doi.org/10.1177/01634437221119021>

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdulah, A. M., Koochang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davidson, R., M, Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M. C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Nord, J. H., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Ribeiro-Navarete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari, M. K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J., Wright, R. (2023). "So What if ChatGPT Wrote It?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy.

International Journal of Information Management, Vol. 71, 1-63.

<https://doi.org/10.1016/j.ijinfomgt.2023.102642>

Fournier, S., & Claudio, A. (2013). Relating badly to brands. *Journal of Consumer Psychology*, Vol. 23(2), 253-264.

<https://doi.org/10.1016/j.jcps.2013.01.004>

Gao, Y., Zhengyu, W., Chen, G. (2018). Alexa, My Love: Analyzing Reviews of Amazon Echo. *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusting Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)*, 372-380.

<https://doi-org.ezproxy.ub.gu.se/10.1109/SmartWorld.2018.00094>

Groh, M., Epstein, Z., Firestone, C., Picard, R. (2021). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences – PNAS*, Vol. 119(1), 1-11.

<https://doi.org/10.1073/pnas.2110013119>

Hochschild, A. R. (1979). Emotion Work, Feeling Rules, and Social Structure. *The American Journal of Sociology*, Vol. 85(3), 551-575.

<https://doi.org/10.1086/227049>

Jantzen, G., & Jensen, J. F. (1993). Powerplay – Power, Violence and Gender in Video Games. *AI & Society*, Vol. 7(4), 368-385.

<https://doi.org/10.1007/BF01891418>

Kayis, A. R., Satici, B., Deniz, M. E., Satici, S. A., Griffiths, M. D. (2002). Fear of COVID-19, loneliness, smartphone addiction, and mental wellbeing among the Turkish general population: a serial mediation model. *Behaviour & Information Technology*, Vol. 41(11), 2484-2496.

<https://doi.org/10.1080/0144929X.2021.1933181>

Kozinets, R.V. (2015). *Netnography: Redefined* (2nd ed.). Sage Publications.

Laranjo, L., Dunn, A. G., Tong, H. L., Ahmet, B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American medical Informatics Association: JAMIA*, Vol. 25(9), 1248-1258.

<https://doi.org/10.1093/jamia/ocy072>

Lee, O. E. K., Nam, I., Chon, Y., Park, A., Choi, N. (2020). Socially Assistive Humanoid Robots: Effects on Depression and Health-Related Quality of Live among Low-Income, Socially Isolated Older Adults in South Korea. *Journal of Applied Gerontology*, Vol. 42(3), 367-375.

<https://doi.org/10.1177/07334648221138283>

Marriot, V., & Pitardi, H.R. (2021). Alexa, she's not human but... Unveiling the driver of consumers' trust in voice-based artificial intelligence. *Psychology & Marketing*, Vol. 38(4), 626-642.

<https://doi.org/10.1002/mar.21457>

Marriott, V., & Pitardi, H. R. (2023). One is the loneliest number... Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction. *Psychology & Marketing*, Vol. 41(1), 86-101.

<https://doi.org/10.1002/mar.21899>

Paul, C. (2015). *Digital Art*. (3rd ed.). Thames & Hudson.

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*, Vol. 149, 1-14.

<https://doi.org/10.1016/j.ijhcs.2021.102601>

Staniloiu, A., & Markowitsch, H. (2012). Gender differences in violence and aggression – a neurobiological perspective. *Procedia – Social and Behavioral Sciences*, Vol. 33, 1032-1036.

<https://doi.org/10.1016/j.sbspro.2012.01.279>

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., Loggarakis, A. (2020). User Experiences of Social Support from Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research*, Vol. 22(3), 1-10.

<https://doi.org/10.2196/16235>

Takayanagi, K., Kazue, T., Shibata, T. (2014). Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot, PARO. *Frontiers in Aging Neuroscience, Vol. 6*, 1-5.

<https://doi.org/10.3389/fnagi.2014.00257>

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. (1st ed.). Volante.

Thorne, S. L. (2020). Hey Siri, tell me a story: Digital storytelling and AI authorship. *Convergence (London, England), Vol. 26*(4), 808-823.

<https://doi.org/10.1177/1354856520913866>

Tunçalp, D., & Lê, L. R. (2014). (Re)Locating boundaries: A systematic review of online ethnography. *Journal of Organizational Ethnography, Vol. 3*(1), 59-79.

<https://doi.org/10.1108/JOE-11-2012-0048>

Turing, A. (1950). Computing Machinery and Intelligence. *Mind, Vol. 59*(236), 433-460.

<https://www.jstor.org/stable/2251299>

Wood, L. J., Zarak, A., Robins, B., Dautenhahn, K. (2019). Developing Kaspar: A Humanoid Robot for Children with Autism. *International Journal of Social Robotics, Vol. 13*(3), 491-508.

<https://doi.org/10.1007/s12369-019-00563-6>

Young, G. (2023). What would strong AI understand consent to mean, and what are the implications for sexbot rape? *Ai and Ethics, Online*, 1-12.

<https://doi.org/10.1007/s43681-023-00383-5>

Zhou, L., Gao, J., Li, D., Shum, H. (2020). The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. *Computational Linguistics – Association for Computational Linguistics, Vol. 46*(1), 53-93.

https://doi.org/10.1162/coli_a_00368

Zdenek, S. (2007). “Just Roll Your Mouse Over Me”: Designing Virtual Women for Customer Service on the Web. *Technical Communication Quarterly, Vol. 16*(4), 397-430.

<https://doi.org/10.1080/10572250701380766>

5.2 Websites

Bardhan, A. (January 18, 2022). *Men Are Creating AI Girlfriends and Then Verbally Abusing Them*. Futurism.

<https://futurism.com/chatbot-abuse>

Blair, E. M. (February 13, 2024). *Addressing Abuse: Why Programming AI ‘Girlfriends’ to Defend Themselves is Vital*. Medium.

<https://medium.com/bouncin-and-behavin-blogs/addressing-abuse-why-programming-ai-girlfriends-to-defend-themselves-is-vital-0acf702e1450>

Chayka, K. (November 13, 2023). *Your A.I Companion Will Support You No Matter What*. The New Yorker.
<https://www.newyorker.com/culture/infinite-scroll/your-ai-companion-will-support-you-no-matter-what>

Delouya, S. (March 4, 2023). *Replika users say they fell in love with their AI chatbots, until a software update made them seem less human*. Business Insider.
<https://www.businessinsider.com/replika-chatbot-users-dont-like-nsfw-sexual-content-bans-2023-2>

Frankze, A. S., Bechmann, A., Zimmer, M., Ess, C. M. and the Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Available at:
<https://aoir.org/ethics/>

Miller, G. (July 17, 2014). *The Moral Hazards and Legal Conundrums of Our Robot-Filled Future*. WIRED.
<https://www.wired.com/2014/07/moral-legal-hazards-robot-future/>

Reddit. (2022). *What do you think about “dark” interactions with Replikas?* Reddit.
https://www.reddit.com/r/replika/comments/xslkxq/what_do_you_think_about_dark_interactions_with/

Reddit. (2024a). *About Reddit*. Reddit.
<https://www.redditinc.com>

Reddit. (2024b). *r/Replika*. Reddit.
<https://www.reddit.com/r/replika/>

Replika. (2024a). *Meet Replika and Create your Replika*. Replika.
<https://replika.com>

Replika. (2024b). *What is Replika?* Replika.
<https://help.replika.com/hc/en-us/articles/115001070951-What-is-Replika>

Replika. (2024c). *How do I change my relationship status with Replika?* Replika.
<https://help.replika.com/hc/en-us/articles/360046490131-How-do-I-change-my-relationship-status-with-Replika>

Replika. (2024d). *What is Replika PRO?* Replika.
<https://help.replika.com/hc/en-us/articles/360032500052-What-is-Replika-Pro>

Similarweb. (2024). *Replika.ai*. Similarweb.
<https://www.similarweb.com/website/replika.ai/#overview>

Schoos, J. (February 2, 2021). *Replika: The AI Companion Who Cares*. Medium.
<https://julia-schoos.medium.com/replika-the-ai-companion-who-cares-e6de52c9a276>

Taylor, A. (January 19, 2022). *Men are creating AI girlfriends, verbally abusing them, and bragging about it on Reddit*. Fortune.

<https://fortune.com/2022/01/19/chatbots-ai-girlfriends-verbal-abuse-reddit/>

Tong, A. (March 21, 2023). *What happens when your AI stops loving you back?* Reuters.

<https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/>

UNESCO. (April 21, 2023). *Artificial Intelligence: examples of ethical dilemmas*. UNESCO.

<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases>