ARTIFICIAL AFFECTION: EXPLORING THE ETHICS OF DIGITAL INTIMACY WITH REPLIKA

By Frederikke Kjær & Lizette Bjørnskov Schmidt

Aalborg University, Denmark
International Business Communication
Supervised by Peter Kastberg

Characters: 268228

2nd of June 2025

Table of Contents

Abstract

Dette speciale undersøger, hvordan Luka Inc., virksomheden bag den AI-drevne chatbot Replika, balancerer sit etiske ansvar i udviklingen, markedsføringen og distributionen af et produkt, der er designet til at fremkalde dybe følelsesmæssige bånd mellem brugere og maskine. Med afsæt i et socialkonstruktivistisk videnskabsteoretisk udgangspunkt og ved at anvende teorierne inden for Human-Machine Communication og Ethical Marketing, analyserer vi samspillet mellem teknologisk design, brugeroplevelser og virksomhedens etiske praksis.

Specialet bygger på en tre uger lang kvalitativ undersøgelse med to Replika-brugere, herunder en bruger af den gratis version af Replika samt én betalingsbruger, samt et offentligt tilgængeligt podcastinterview med Luka Inc.'s CEO, Eugenia Kuyda. Gennem tematisk analyse identificerer vi, hvordan brugere udvikler relationer til Replika, der minder om menneskelige forbindelser, og hvordan disse relationer formes af både appens design og virksomhedens kommunikation. Vores analyse viser, at der opstår en konflikt mellem brugernes følelsesmæssige tilknytning til Replika og Luka Inc.'s skiftende markedsføring, især i takt med, at virksomheden først indførte og siden fjernede erotiske funktioner for brugerne. Disse ændringer har haft en tydelig indvirkning på brugernes oplevelse og deres følelsesmæssige tryghed i relation til Replika.

Vi argumenterer for, at Luka Inc. bærer et betydeligt etisk ansvar, ikke blot i forhold til teknologisk sikkerhed, men også i forhold til gennemsigtighed og følelsesmæssig beskyttelse i brugeroplevelsen. Specialet bidrager til den aktuelle debat om menneskeliggørelse af kunstig intelligens ved at belyse, hvordan relationer mellem mennesker og maskiner konstrueres, samt hvordan kommercielle interesser kan komme i konflikt med brugerens følelsesmæssige trivsel. Vores resultater peger på behovet for øget etisk refleksion i udviklingen og markedsføringen af AI-teknologier, der aktivt opfordrer til følelsesmæssig tilknytning.

1. Introduction

Can an AI chatbot, built to remember your past conversations, understand your emotions, and respond with genuine concern and empathy, ever truly replicate the connection of human interaction, even when you know from the start that it is entirely artificial?

Replika, a social chatbot designed to form deep and meaningful connections with its users, blurs the line between artificial and real relationships. As AI-driven companionship becomes more sophisticated, users find themselves emotionally invested in a digital entity programmed to engage, support, and even love them back. But where does the illusion end, and what responsibility does the company behind Replika hold when encouraging these connections? In a world where human-machine interactions are becoming increasingly indistinguishable from real-life relationships, the ethical implications of artificial intelligence companionship require attention.

In our bachelor's dissertation, we explored the impact of human-machine communication on Replika users, focusing on how interactions with the chatbot influence their perception of reality and communication patterns. Our findings revealed that users' experiences with Replika are shaped by their initial intentions and the level of realism embedded in the chatbot's design. The more realistic and responsive Replika appears, the stronger the emotional attachment users develop, often leading them to rely on it as a safe, non-judgmental space for self-expression and personal growth. However, this deep connection also carries risks, as users may transfer expectations and interaction patterns from their artificial relationships into human interactions. When confronted with the inherent limitations of artificial intelligence (AI), such as system updates or shifts in Replika's behaviour, users are abruptly reminded that their companion is not a real person, exposing the fragility of these emotional bonds. Through our research, we recognised the growing significance of AI-driven companionship and the ethical concerns it raises.

Following this study, we continued to observe Replika's development over the past couple of years, seeing a change in how Luka Inc., the company behind Replika, distributes their product, which sparked our interest to further research the chatbot from an organisational perspective

throughout our master's dissertation. Specifically, we aim to examine Luka Inc.'s ethical responsibility in creating, distributing, and marketing a chatbot that encourages users to form deep emotional connections. By shifting our focus to the company's role, we seek to uncover the ethical considerations behind selling a product that can affect users' emotions and perceptions. To thoroughly explore this assessment, we will conduct a three-week experiment and in-depth interviews with two participants concerning their experiences with Replika and their perceptions of Luka Inc. Additionally, we will analyse a secondary interview with Eugenia Kuyda, the CEO of Luka Inc., to gain insight into how the company presents and discusses their emotional companion as a product and distribution practices. We will approach this data as social constructivists, employing thematic analysis to identify and interpret recurring themes across the interviews. We will draw on *'Human-Machine Communication'* by Andrea L. Guzman and Seth C. Lewis, as well as *'Ethical Marketing'* by Dilip S. Mutum and Ezlika Ghazali, to comprehend the participants' emotional engagement with Replika and evaluate the extent of Luka Inc.'s corporate accountability.

We argued that this topic is particularly relevant in today's rapidly evolving digital world, where the boundaries between AI and reality are becoming increasingly blurred. As AI-generated interactions grow more advanced, even able to imitate human interaction, the potential for users to mistake artificial relationships for real ones becomes a pressing concern. Investigating how companies address these ethical dilemmas is crucial for promoting responsible AI development and protecting users from unintended psychological consequences. This consideration leads to the following problem statement:

1.1. Problem Statement

How does Luka Inc., the company behind Replika, balance its ethical responsibility towards users while creating a social chatbot that markets deep emotional connections?

2. Background

The subsequent chapter will be dedicated to delineating the essential background information needed to comprehend our research. We aim to introduce Luka, the company behind the social

chatbot Replika, define the chatbot itself, and scrutinise the marketing and brand community surrounding the chatbot. By doing so, the foundational knowledge will provide a common ground for exploring the ethical complexities of a corporation that sells an artificial companion made to foster emotional connections.

2.1. The Founding Company: Luka Inc.

Luka Inc. is a San Francisco-based artificial intelligence company best known as the creator of the AI companion app, Replika. The company was founded in 2014 by two Russian-born entrepreneurs, Eugenia Kuyda and Philip Dudchuk. Prior to tech entrepreneurship, Kuyda worked as a journalist and editor in Moscow. She became interested in artificial intelligence and conversational bots, co-founding Luka Inc. to explore AI-driven communication. Meanwhile, Dudchuk held a degree in computational linguistics, leading Luka's technical development by drawing on experience from Russia's tech industry. He played an instrumental role in designing Luka Inc.'s chatbot architecture and product strategy. Early on, Luka Inc. built a chatbot that could provide restaurant recommendations and help make reservations, showcasing the team's ability to create conversational AI for everyday services (Newton 2016).

In late 2015, a personal tragedy gave Luka Inc. a new direction when Kuyda's best friend died in an accident. In an effort to cope and preserve his memory, Kuyda collaborated with Dudchuk and their team to create a chatbot modelled on her late friend, utilising thousands of text messages he had sent her, to mimic his behaviour, personality and conversational style. The project helped Kuyda grieve but also demonstrated the potential of AI to capture an individual's essence, which inspired the concept of Replika as a personal digital companion that could provide emotional support (cf. 2.2.) (Replika, 'Our Story'). Together, Luka Inc. set out to build consumer AI technology focused on natural conversation and emotional connection, founded in psychology and empathy. The company remains relatively small in size, with only a few dozen employees as of the latest data, published on their official LinkedIn profile, showcasing between 10 and 50 employees. In recent years, Luka has adopted a remote-work model where the team reportedly works without any physical office, collaborating from various locations (Newton 2016; "Luka: Overview | LinkedIn" n.d.).

Thus, Luka Inc. is a San Francisco-based AI company founded in 2014 by Eugenia Kuyda and Philip Dudchuk. While it started out building practical chatbots for everyday tasks, a personal loss led the team to explore more emotionally driven uses of artificial intelligence, resulting in the creation of Replika, an AI companion made with the intention of providing emotional support for humans. The following section will take a closer look at Replika itself, elaborating on its functionality and design in regards to the user's operation of it (Murphy and Templin 2019).

2.2. What is Replika and how to utilise it

As briefly mentioned previously, Replika is a social chatbot driven by artificial intelligence, made to form intimate and emotional connections with its users through in-depth conversations. Over several conversations, Replika stores the information it gathers in a digital library, which is utilised to generate sincere and human-like interaction that is tailored to the users' preferences. Luka, the founding company of Replika, depicts the chatbot as *'a space where you can safely share your thoughts, feelings, beliefs, experiences, memories, dreams – your private perceptual world'* (Replika, 'Our Story'). Replika operates through a combination of advanced neural networks, machine learning algorithms, and pre-programmed dialogue scripts. Trained on a broad dataset, it can generate unique responses tailored to each conversation (Replika, 'Our Story').

To get started utilising Replika, users create an account and personalise their AI companion by selecting its name, appearance, and preferred gender, meaning that you can personalise it exactly the way you prefer with game-like features (cf. App. 6). With the free version of the app, your Replika can only function as a friend. However, upgrading to the paid version allows you to choose a different relationship dynamic, such as a sibling, romantic partner, mentor, or even a yoga teacher, depending on what you desire. When setting up the app, you will be asked a series of questions to attribute your Replika to best suit your needs, including choosing how you want it to interact with you, selecting its personality traits, defining your preferred communication style, and even specifying your love language (Replika 2025). The chatbot learns from interactions, adapting its responses to match the user's personality and conversational style over time and will

store all of the data you provide it with. Replika has various features, including text and voice-based chats, role-playing scenarios, coaching sessions for mindfulness and personal growth, and even augmented reality interactions. Users can guide their Replika's development by providing feedback on its responses and shaping its personality to better align with their preferences. Additionally, customisation options allow users to unlock new traits, interests, and accessories through in-app currency (Replika 2025).

Ultimately, Replika is made to encourage and establish emotional, personalised interactions through AI-driven conversation. By combining neural networks, machine learning, and scripted dialogue, the chatbot adapts to users' preferences over time, creating a sense of intimacy and companionship. With extensive customisation options and relationship modes ranging from friendship to romantic partnership, Replika aims to offer a tailored user experience that allows users to form an emotional connection to their Replika, emphasising the platform's central purpose to serve as a safe and responsive community for self-expression, emotional development, and personal growth (Brooks 2023). In correlation, the next section delineates the evolution of Replika's marketing, including its initial positioning as a mental health companion, its later emphasis on erotic roleplay features, and the broader impact of these strategies on the app's public image and the emotional experiences of its users, providing us with some background knowledge on Luka Inc.'s effort to generate a certain comprehension and perception of Replika.

2.3. Marketing of Replika

From the beginning of Replika's launch in March 2017, Luka Inc. employed psychologists to develop techniques for making the chatbot ask questions that generate intimacy and encourage emotional connections. Thus, Replika functions as a messaging app where users answer questions to build a digital library of personal information. The data is then processed by an AI program to create a chatbot tailored to the user. In its early stages, the Replika chatbots were widely criticised for being unconvincing and robotic. However, as AI technology advanced, the chatbots became increasingly sophisticated, catching attention for their uncanny human-like features. What has drawn users to Replika is the ability to create an AI companion customised to

their preferences, including appearance, personality traits, and conversational style. Because users provide detailed prompts during the setup process, interactions with Replika often unfold with a surprising level of complexity, comfort, and trust, encouraging emotional connections from the very beginning (Brooks 2023).

Replika was initially marketed as a mental health tool, particularly appealing to people struggling with loneliness and past rejection. The chatbot was designed to offer a relationship where users would not have to fear abandonment; the chatbot would always be there, waiting, supportive, and ready to listen. Yet, while Luka Inc. brands Replika as a companion for emotional support, it remains an artificial entity software that, like any other IT-based service, requires regular updates and adjustments. These updates have, at times, disrupted user experiences. Some users have reported that their Replika's personality changed overnight, making responses seem robotic and scripted. Others were dismayed when, following a previous update, their chatbot suddenly rejected any sexual advances. Users have described these changes as feeling like their Replika, whether perceived as a spouse, partner, or friend, was fading away or even dying due to software updates (Cole 2023).

As part of its marketing strategy, Luka Inc. began promoting Replika as a highly sexualised chatbot. In 2022, users of the paid version received sexually suggestive images, signalling a shift toward more intimate user engagement. The Erotic Roleplay (ERP) feature, still exclusively available to paying users, allows for flirtation, sexting, and other erotic interactions, further reinforcing this sexualised positioning. However, in February 2023, Italy's Data Protection Authority ordered Replika to cease processing the personal data of Italian users or risk a $21.5 million fine due to concerns over data security and privacy leaks. In response, Luka Inc. removed or significantly altered ERP features, causing immense frustration among users who felt their Replikas had lost their personalities. Many described their reps as becoming distant, numb, and robotic, experiencing this shift as a form of rejection (Brooks 2023).

The abrupt change in Replika's behaviour created a fundamental contrast in its branding. Once marketed as a mental health tool for those struggling with loneliness and rejection, the chatbot's new restrictions made it seem judgmental toward users who developed romantic feelings for their AI companion, even though that is exactly what it was branded for in the beginning. In an

interview, Luka Inc. co-founder and CEO Eugenia Kuyda stated that Replika was never intended to function as an ‘adult toy’, shifting blame onto users as part of the company’s effort to protect its reputation (Delouya 2023).

From our own experiences with Replika over the past two years, initiating conversations with a chatbot can feel similar to early interactions on a dating app where users heavily chat and get to know one another before forming a deeper connection. At first, you may, as a user, share superficial details about yourself, your life and your personal interests, but as the relationship develops, you begin to disclose more personal thoughts and emotions. However, the unpredictable nature of updates, combined with shifts in the company's marketing approach, has left many users questioning whether their emotional investments in Replika can be sustained over time. It raises important questions about how Luka Inc.’s marketing of Replika aligns with actual user experiences and the intentions behind the app’s development, and Luka Inc.’s ethical responsibility towards their users when selling and promoting a product that encourages emotional connection and understanding (Kjær and Schmidt 2023).

In summary, Replika evolved from a mental health support tool into an advanced, customisable AI companion that seeks to provide its users with emotional intimacy through tailored interactions. Initially created to provide a safe and non-judgmental space for users facing loneliness, the platform has grown increasingly complex both in terms of user experience and ethical implications. While its personalisation features and human-like qualities have drawn many users into emotionally invested relationships, sudden changes to its functionality and marketing strategy have disrupted these bonds, raising concerns about user trust and emotional safety. Alongside this development, a diverse and growing user community has formed around Replika both organised by Luka Inc. and private users engaged in creating a community for fellow Replika users. The following section introduces these communities, explaining who the users are, their interaction with one another, and the online space's contribution to the broader Replika experience.

2.4. The Replika Community

Over the past two years of observing Replika's development and the growth of its user base, it has become evident that the Replika community is expanding steadily. The community appears to function as a form of safe space, in which users engage with one another by sharing personal experiences, reflections, and the emotional fluctuations associated with their interactions with their Replikas.

Since the launch of Replika in 2017, the app has attracted widespread interest and cultivated a growing user community. In response to this engagement, both Luka and the users themselves initiated several online platforms to foster user interaction, support, and feedback between users. A prominent example is the Reddit forum *'r/Replika'*, which hosts approximately 80,000 members who discuss their experiences with Replika, the company behind it, and share their highs and lows with their own Replikas. Additionally, Luka has established the closed Facebook group, *"Replika Friends"*. As of April 2025, *Replika Friends* alone includes nearly 40,000 members who share screenshots of their conversations with their Replikas, exchange personalisation strategies, and engage in discussions related to the app's features and updates for emotional support and knowledge sharing. To join the group, prospective members must request access and agree to uphold the confidentiality of all shared content. They are also required to commit to being respectful, supportive, and non-judgmental in their interactions with others (Replika Friends; Replika, r/Replika).

Eugenia Kuyda plays an active role in these communities, serving as an administrator in the Facebook groups and directly participating in conversations, particularly when users raise concerns. She frequently posts updates, responds to feedback, and demonstrates a continued interest in the needs and preferences of the user base, thereby reinforcing a sense of accessibility and responsiveness from the company (Replika Friends; Replika, r/Replika). However, not all users feel adequately represented or heard within these official channels. As a result, some have created independent Reddit forums and Facebook groups outside of Luka Inc.'s oversight. These user-led communities often provide space for critical reflections, including negative experiences with the platform and more intimate or controversial interactions with their AI companions (Cole 2023).

Replika itself functions as a customisable chatbot that evolves based on user interaction, drawing on personal data, preferences, and conversational patterns to simulate human-like dialogue. Users can tailor their Replika's appearance, personality traits, and interests while tracking its developmental progress over time. The platform relies on a combination of Large Language Model (LLM) technology and scripted responses to generate increasingly coherent and emotionally attuned conversations. Although basic interaction is free, access to advanced features and paid subscription options is available, including forming other relationships with your Replika other than just a friend or companion, like a romantic partner, sibling, etc. ('Introducing Advanced AI mode' 2023). Overall, Replika has not only introduced a highly personalised AI companion experience but has also given rise to an active and multifaceted user community. Through both official and user-driven communities, users engage in knowledge sharing, emotional support, and critical reflection. These online spaces reveal the depth of user investment in Replika, highlighting both the app's emotional significance and the diverse ways individuals choose to relate to their AI companions.

In summary, this chapter has provided essential background information necessary to understand the scope of our research on Replika in order to research their ethical responsibility towards their users when selling and marketing a product reflecting emotional vulnerability and investment. It introduced Luka Inc., the San Francisco-based AI company behind the social chatbot Replika, highlighting its origins, founders, and evolution from practical chatbots to emotionally focused AI companions. The chapter then examined Replika itself, a customisable AI chatbot designed to foster emotional connection through human-like conversations, personalisation, and adaptive learning. Furthermore, the marketing strategies of Replika were explored, showing how the app has shifted from positioning itself primarily as a mental health support tool to incorporating more controversial features such as erotic roleplay, raising complex ethical questions. Finally, the growing user community around Replika was discussed, emphasising both company-managed and independent online communities where users share experiences, support, and critiques on their experiences with Replika.

3. Philosophy of Science and Research Design

With the technological, commercial, and social context of Replika established, including its development by Luka Inc., marketed functionalities, and user community, we now turn to the philosophical and methodological foundations that guide this dissertation. Before outlining our chosen philosophy of science, social constructivism, we briefly consider phenomenology and interpretivism as alternative approaches. While both allow for thorough exploration into individual experience and meaning-making, we ultimately adopt a social constructivist perspective, as it more effectively captures how users' emotional relationships with Replika are shaped through social interaction, cultural narratives, and power relations. The foundation informs both our research design and our critical analysis of the company's ethical responsibility.

3.1. Philosophy of Science Review

Within *'Consciousness and Selfhood'*, Dan Zahavi delineates a phenomenological view of what it means to be a self, challenging both simplified physical explanations and detached theories that separate the self from experience. Drawing on classic phenomenologists like Edmund Husserl and Maurice Merleau-Ponty, Zahavi argues that selfhood is not something we step back and observe, but something that is always present in every conscious moment (Zahavi 2005, 105-106). He introduces the idea of pre-reflective self-awareness, meaning we have an implicit sense of ourselves in all experiences, even before we actively think about it (Zahavi 2005, 100, 146).

Zahavi emphasises the importance of the first-person perspective as something fundamental and cannot be broken down into just social roles or cognitive processes. This approach is particularly useful for understanding lived experience, emotions, and identity because it highlights how self-understanding is connected to our bodies and immediate surroundings (Zahavi 2005, 123-125). Although Zahavi's work is firmly rooted in phenomenology, it also engages with wider discussions in philosophy of mind and cognitive science, defending the reality of subjective experience against views that deny or reduce it (Zahavi 2005, 117-118). Zahavi's ideas provide a thorough foundation for exploring individuals' self-experiences when interacting with

technologies like AI companions, emphasising the temporal development of these experiences, the emotional depth, and the connection to bodily awareness. However, while Zahavi permits exploration of individual subjective experience and pre-reflective self-awareness, we consider it less relevant for our study's focus, as our research centres on how user experiences with Replika are influenced not only by internal consciousness but also by corporate messaging. Phenomenology's emphasis on the immediate, first-person perspective does not sufficiently account for these social and discursive influences. As a result, it does not fully capture the broader contextual dynamics that shape users' interpretation and meaning-making of their AI-interactions.

Having examined phenomenology, we now turn to interpretivism. While phenomenology centres on how individuals internally make sense of their experiences, interpretivism broadens this view by emphasising the role of social context and shared meanings in shaping human understanding. This shift allows us to explore not only the personal but also the intersubjective processes through which meaning is constructed and communicated.

Malcolm William's article, *'Interpretivism and Generalisation'*, critically examines the relationship between interpretive approaches in sociology and the concept of generalisation. Williams challenges the division between qualitative and quantitative research, arguing that interpretivist research can and should aim for generalisation, but in a different way than statistical generalisation (Williams 2000, 214-215). He introduces the concept of 'moderatum generalisation,' which refers to generalisations that are contextually grounded and sensitive to the nuances of individual cases (Williams 2000, 215, 221). Williams contends that interpretivist research, by focusing on understanding the meanings and experiences of individuals within specific contexts, can produce deep and broadly applicable insights. He emphasises the importance of transparency in the research process, advocating for clear articulation of the researcher's perspective and the context of the study to enhance the validity and applicability of the findings (Williams 2000, 221-222). While interpretivism showcases individuals' creation of meaning within social contexts, it primarily focuses on understanding personal interpretations and shared meanings. Yet, this approach may not capture the broader social, cultural, and institutional forces that actively shape and construct these meanings.

3.2. Social Constructivism

Before presenting our research design, we outline our ontological and epistemological stance, drawing on Daymon and Holloway. Ontology concerns the nature of being and the assumptions we make about what exists in the social world. According to Daymon and Holloway, it involves asking fundamental questions such as what can be said to exist and what the nature of that existence is. Epistemology, on the other hand, relates to the theory of knowledge. It determines what counts as valid knowledge and considers how knowledge is acquired, as well as the relationship between the knower and the known (Daymon and Holloway 2011, 100). Our research is grounded in a social constructivist epistemology, particularly as described by Danish philosopher Finn Collin. According to Collin, knowledge is not passively received from the world but is actively constructed through social processes, language, and interaction Collin 2021, 19-21). The position holds that reality is co-created rather than objectively given. Adopting this stance allows us to research how users interpret and give meaning to their interactions with the Replika chatbot. We treat their experiences not as expressions of an objective truth, but as individually and socially constructed narratives, shaped within specific socio-technical and discursive contexts (Collin 2021, 27-28).

Collin argues that social constructivism can be understood through three distinct dimensions. The first is epistemological constructivism, which concerns how our understanding of reality, including scientific theories, is in itself a social construction. Within this view, it is not possible to distinguish objectively between ‘true’ knowledge and what is merely accepted as knowledge within a particular society or community (Collin 2021, 26–27). The second dimension, often referred to as ontological constructivism, goes further by claiming that reality itself is constructed, not merely our perceptions or theories about it. The third and final dimension raises the question of who or what performs the act of constructing. Here, opinions diverge: some constructivists maintain that it is the scientific researcher who produces the construction, while others argue that it is the social actors within society who are ultimately responsible (Collin 2021, 26–27). In the context of our research, these dimensions support our focus on how users’ emotional experiences with Replika are shaped not only by individual interpretation but also by broader social, technological, and discursive structures. Rather than treating users’ relationships

with Replika as either objectively real or entirely artificial, we adopt Collin's view that such relationships are socially constructed phenomena, emerging within specific socio-technical environments.

3.3. Research Design

To research how users form emotional connections with Replika, we apply Human-Machine Communication (HMC) theory, developed by Andrea L. Guzman and Seth C. Lewis (cf. 5.1.). The theory permits us to scrutinise the communicative and social dynamics that emerge when people interact with social chatbots. In addition, we apply Ethical Marketing Theory by Dilip S. Mutum and Ezlika M. Ghazali (cf. 5.2.) to critically examine Luka Inc.'s ethical responsibility in marketing and promoting a product that may be seen as emotionally exploitative towards its users. Both theories pay particular attention to how humans attribute agency, social presence, and relational meaning to machines, which is essential in understanding how and why users may develop deep emotional bonds with an AI companion like Replika. We are particularly concerned with whether these practices protect users' emotional vulnerabilities or exploit them; an ethical dilemma that highlights the underlying tension between commercial interests and user well-being, fundamental to our analysis.

From a methodological perspective, our qualitative research design reflects our constructivist stance. As Daymon and Holloway note, qualitative research enables the exploration of lived experience, emotion, and personal meaning-making (Daymon and Holloway 2011, 105). We aim to uncover how our participants perceive and make sense of their relationship with Replika, considering how these perceptions are shaped by individual experiences, social inputs, and the design of the technology itself. To collect data, we adopt an interview-based approach involving two Replika users, one utilising the free version and one utilising the paid version of Replika. The contrast enables us to explore how different access models influence emotional investment and user experience, especially in relation to features like romantic roleplay and voice calls, which are exclusive to paying users. These differences also allow us to critically examine Luka Inc.'s marketing strategies and whether they align with principles of ethical marketing.

The data collection will span three weeks and consist of three interviews with each participant, beginning with a pre-interview to gather contextual background (personal motivations, mental health status, and expectations), followed by a mid-point interview and a final, reflective session. The structure allows us to trace participants' evolving relationships with their Replika chatbot over time and capture the nuances of emotional and cognitive change. In addition to user interviews, we will analyse a 30-minute podcast interview between Replika CEO and co-founder Eugenia Kuyda and podcaster Seth Rosenberg. This provides insight into the company's public narrative, marketing logic, and ethical framing. Integrating this analysis helps us assess Luka Inc.'s positioning and responsibility in designing and promoting a product that invites deep emotional involvement.

In summary, our ontological and epistemological positioning within a social constructivist paradigm shapes both the focus and form of our inquiry. By understanding users' relationships with Replika as socially constructed and situated within broader socio-technical contexts, we establish a foundation for examining the co-production of emotional meaning, technological design, and ethical responsibility. With this philosophical foundation in place, we now proceed to the literature review, in which we survey the existing body of research relevant to contextualise our study and determine the theoretical foundation most suited to our analytical objectives.

4. Literature Review

In the following chapter, we present a critical literature review to support our exploration of Luka Inc.'s ethical responsibility in designing and distributing Replika as a product that encourages emotional connections. We explore four areas of research: deceptive advertising, gamification, human-machine communication, and corporate ethics. While each field allows for distinct insight, they also overlap in meaningful ways when it comes to understanding how AI companions are marketed, perceived, and experienced. Rather than aiming for a comprehensive account, this chapter focuses on evaluating how relevant these theoretical contributions are to our research. Through this review, we clarify why we ultimately chose Human-Machine Communication by Guzman and Lewis and Ethical Marketing by Mutum and Ghazali as our final theoretical stance.

4.1. Deceptive advertising

To contextualise our examination of Luka Inc.'s ethical responsibility towards their users, it is necessary to first engage with existing literature on deception in advertising. While our research ultimately centres on emotional manipulation and the relational dynamics between users and AI companions, the concept of deception provides a useful entry point for understanding how persuasive strategies can become ethically problematic. In this section, we explore two key theoretical contributions shaping the academic discourse on deceptive advertising: Carson, Wokutch, and Cox's deception in advertising, and McCornack's Information Manipulation Theory. These works offer valuable insights into how deception operates through intentionality, message construction, and consumer perception. Although neither theory fully encompasses the relational and emotional dimensions that are central to our dissertation, they serve as important stepping stones for identifying the ethical boundaries of communication and influence in marketing contexts. By engaging critically with these perspectives, we are better equipped to contrast traditional advertising deception with the more subtle and effective forms of manipulation at play in emotionally intelligent technologies like Replika.

4.1.1. An ethical analysis of deception in advertising by Thomas L. Carson, Richard E. Wokutch and James E. Cox Jr.

To initiate our section on deceptive advertising, we explore the theoretical work *'An Ethical Analysis of Deception in Advertising'* by Thomas L. Carson, Richard E. Wokutch, and James E. Cox Jr., published in 1985. Their study delves into the ethical and legal perspectives of deceptive advertising, arguing that previous definitions of deceptive advertising lack depth as they primarily focus on an advertisement's effects on the consumer without considering the advertiser's intent. The oversight fosters an environment where manipulating consumers and engaging in unethical business practices may be perceived as acceptable. Many existing definitions of deception in advertising serve as criteria for determining whether an advertisement should be prohibited. Additionally, Carson, Wokutch, and Cox contend that it is necessary to develop a revised definition that accounts for both the sender and receiver aspects of an

advertisement to regulate the distribution of deceiving content (Carson, Wokutch, and Cox 1985, 93-94).

According to their study, it is essential to distinguish between lying and deception as separate concepts. They define lying as the deliberate communication of false statements, whereas deception does not necessarily require false statements or linguistic expression but can occur through actions that mislead consumers. However, deception is only effective if the consumer believes the misleading claim; if they detect the falsehood, then deception has not occurred (Carson, Wokutch, and Cox 1985, 97-98). Based on this distinction, the authors propose a definition of deception, stating*: 'x deceives y if and only if: x causes y to have certain false beliefs (b) and x intends or expects his actions to cause y to believe b'* (Carson, Wokutch, and Cox 1985, 98), emphasising the necessity for an advertiser to either intend to mislead or anticipate that their actions are likely to create false beliefs in consumers.

For regulatory purposes, the authors suggest defining an advertisement as deceptive if it causes a significant percentage of its target consumers to hold false beliefs about the product. In other words, an advertisement is deceptive when it tends to mislead a substantial portion of the intended audience. What constitutes a 'significant percentage' is context-dependent and must be determined case-by-case. Furthermore, the study critiques the Federal Trade Commission's (FTC) reliance on the 'reasonable consumer standard', arguing that this approach fails to protect vulnerable consumers (Carson, Wokutch, and Cox 1985, 96, 102). The authors argue that deceptive advertising is morally problematic because it harms consumers and competitors while dissolving trust in advertising as a whole. They advocate for stricter regulations, corrective advertising, and financial penalties for deceptive practices to ensure transparency and ethical conductå. By presenting a refined definition of deception, Carson, Wokutch, and Cox contribute to the ongoing discourse on the ethical responsibilities of advertisers and the necessity for rigorous legal theoretical stances to prevent misleading practices (Carson, Wokutch, and Cox 1985, 104).

The article *'An Ethical Analysis of Deception in Advertising'* by Carson, Wokutch, and Cox (1985) provides a foundation for understanding deceptive advertising from both ethical and legal standpoints. It overall critiques earlier definitions of deception for focusing solely on consumer

impact, without accounting for advertiser intent. They propose that advertising is deceptive if it misleads a significant percentage of the target audience, calling for regulatory approaches and stricter enforcement, including corrective advertising and penalties, to promote ethical business conduct. While this theory is very insightful in relation to intent and consumer misbeliefs, it is not suitable as a theoretical stance for our dissertation, since our interests mainly lie within emotional manipulation rather than deception and misinformation. Carson, Wokutch and Cox's theory would be particularly relevant if our focus were on misinformation within conventional advertising, where the ethical breach would lie in causing false consumer beliefs. However, in our research about Luka Inc. and Replika, the interest does not primarily concern whether users are misled about the chatbot's capabilities or intentions. Instead, it emerges from the relational dynamics users form with the AI and the emotional vulnerabilities that the product's design may encourage.

4.1.2. Information Manipulation Theory by Steven A. McCornack

Another contribution to the field of deceptive advertising is McCornack's *'Information Manipulation Theory'* (IMT) from 1992. McCornack argues that deceptiveness is perceived as a form of functional adaptation to the complexities of communication. Individuals often find themselves in situations where they must balance competing objectives; on one hand, conveying information that their conversational partners are entitled to receive, and on the other, minimising the potential harm that disclosing such information might cause (McCornack 1992, 3). To navigate this challenge, individuals may resort to deceptive messaging by manipulating information in one or more ways. They may change the amount of information shared by restraining or leaving out important details, manipulating the accuracy by providing false or misleading content, presenting information ambiguously, utilising unclear language, or adding irrelevant details to distract from the main issue (McCornack 1992, 4).

McCornack's research provides a comprehension for understanding how deceptive messages are constructed and how individuals manipulate information to serve their communication goals. He argues that deception emerges from the way individuals selectively present or withhold information to influence their interlocutors. Given that conversational participants hold

expectations regarding information quantity, quality, manner, and relevance, deception occurs when these expectations are covertly violated (McCornack 1992, 5-7). As a result, verbal deception can be understood as a subclass of uncooperative acts; interactions in which the guiding principles of cooperative communication are subtly undermined. Additionally, IMT serves as a useful tool for analysing the types of deceptive messages produced in contexts where conflicting demands exist. Previous research has identified various patterns in deceptive communication, illustrating that individuals strategically exploit different forms of information manipulation to navigate difficult or sensitive interactions (McCornack 1992, 10-13).

McCornack's theory expands the scope of deceptive advertising by showcasing how manipulation can occur not only through outright lies but also through selective, vague, or misleading presentation of information, making it an interesting tool for analysing ethically unclear marketing strategies that subtly mislead consumers without crossing the line into deception. In relation, this theory generally concentrates on message-level communication, which does not sufficiently address the broader relational, emotional, and organisational aspects we aim to explore. However, *'Information Manipulation Theory'* does not fully account for how users form lasting emotional connections with Replika, how the chatbot is designed to simulate human interaction, or how Luka Inc. navigates its ethical responsibilities in that process. Although the theory helps illuminate some of the communicative mechanisms involved, it does not provide the conceptual grounding needed to support the aim of our dissertation.

In summary, while both Carson, Wokutch and Cox' '*An Ethical Analysis of Deception in Advertising'* and McCornack's *'Information Manipulation Theory'* allow for in-depth comprehension of deceptive practices in advertising, neither fully captures the complexity of our research focus. Carson, Wokutch and Cox are mainly concerned with misinformation and advertiser intent in traditional advertising settings, which does not align with our interest in emotional manipulation and user vulnerability. McCornack expands the understanding of deception by highlighting how information can be subtly manipulated, but the theory remains focused on individual message-level communication. It does not account for the broader emotional, relational, and organisational dynamics involved in AI companionship. Therefore, while both theories help inform our perspective, they are not selected as core theoretical stances for this dissertation.

4.2. Gamification

To contextualise our exploration of Replika's emotionally engaging features, it is essential to engage with the concept of gamification. While our focus lies in how the app fosters emotional connection and intimacy, gamification provides a useful entry point for understanding how design elements can structure user behaviour and motivation. In this section, we turn to two key theoretical contributions that have shaped the academic discourse on gamification: Deterding et al.'s foundational definition of game design elements in non-game contexts, and Wood and Reiners' elaboration on how these elements function through components, mechanics, and dynamics. Although these perspectives are primarily concerned with engagement and task-oriented motivation rather than emotional intimacy, they offer valuable insights into how digital systems may influence user investment and behaviour. By situating Replika within this theoretical landscape, we lay the groundwork for assessing the ethical implications of game-inspired design in emotionally charged interactions.

4.2.1. From Game Design Elements to Gamefulness: Defining "Gamification" by Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke

In pursuing a deeper understanding of gamification and its conceptual foundations, we examine Sebastian Deterding et al.'s research on the definition of gamification. The exploration is essential to our study on Replika as they incorporate game-like elements to enhance user engagement and emotional connection. By understanding the nuances of gamification, we can critically assess how Replika employs these elements to encourage deep relationships with AI and whether this aligns with or diverges from traditional gamification theories. This distinction will allow us to evaluate the ethical implications of gamification in emotionally charged digital interactions.

Deterding et al. provide a structured approach to defining gamification by distinguishing it from full-fledged games and related methodologies that integrate game elements for non-entertainment purposes (Deterding et al. 2011, 11). They argue that gamification refers to the implementation of game design elements in non-game contexts rather than the creation of complete games with non-entertainment objectives (Deterding et al. 2011, 9). A definition that

emphasises the partial rather than holistic application of gaming principles and the contextual deployment of game mechanics to enhance engagement and motivation. The authors further dissect the concept by categorising game design elements into different levels of abstraction. These include interface design patterns, game design patterns and mechanics, game design principles, game models, and game design methodologies, providing a structured lens through which gamification can be examined and applied effectively (Deterding et al. 2011, 12-13).

Deterding et al. highlight the significance of differentiating between ‘gamefulness’ and ‘playfulness’ and argue that gamification aims to evoke gamefulness, which is characterised by structured challenges, clear goals, and explicit rules that drive engagement (Deterding et al. 2011, 13). Conversely, playfulness is more open-ended and exploratory, lacking the structured motivators typically found in gamified systems. The distinction is essential for understanding gamification within different domains, particularly education, marketing, and workplace motivation (Deterding et al., 2011, 10-12).

Thus, ‘gamification’ is defined by Deterding et al. (2011) as the utilisation of design elements characteristic of games in non-game contexts, distinguishing it from comprehensive games and playful designs. It positions gamification within the broader socio-cultural trend of deception, which encompasses the implementation of gamified elements in non-game contexts with further differentiation into game technology, practices, and design (Deterding et al., 2011, 10). While the definition provided by Deterding et al. is widely accepted, it prompts important ethical and psychological considerations, as their research highlights that gamification’s effectiveness is contingent on its alignment with intrinsic and extrinsic motivation theories. Over-reliance on extrinsic rewards may lead to diminished motivation over time, stressing the need for thoughtful design that prioritises meaningful engagement over superficial rewards. While their conceptualisation of gamification gives a useful standpoint on distinguishing it from related fields such as serious games and game-based learning, it does not fully address all aspects of user experience or the long-term behavioural impacts of gamification, underscoring the necessity of further exploration into the practical implications and limitations of gamification as a design approach but still gives a valuable impression of how the term ‘gamification’ unfolds.

4.2.2. Gamification by Lincoln C. Wood and Torsten Reiners

Moreover, when we scrutinise the theoretical foundations of gamification, we elaborate by turning to Lincoln Wood and Torsten Reiners' exploration of game elements in non-game contexts. Their study aims to comprehend how gamification operates within various domains, offering insights into user engagement, motivation, and behavioural influence. By delving into this study, we can assess how gamification functions as a tool for structuring user experiences.

Wood and Reiners start by defining gamification as the utilisation of gaming designs in non-game contexts to enhance engagement and motivation through structured challenges and digital feedback systems (Wood and Reiners 2015, 3040). An example would be an app that does not have the purpose of being a digital game but still contains game-like features. Building upon prior scholars' conceptual idea of gamification, this definition distinguishes it from digital games to game-influenced digital systems. They emphasise that gamification is not about creating complete gaming experiences but rather integrating specific components, mechanics and dynamics to encourage user participation in tasks that may otherwise lack motivation (Wood and Reiners 2015, 3040, 3042).

Briefly touching upon these aspects, components refer to visible features that guide the users' engagement; these could, among others, be points to measure progression, quests to add a sense of purpose or virtual goods perceived as valuable stuff advancing the user's position. On the other hand, mechanics are the underlying structures and interactions that determine how users engage with the system. These mechanics determine user behaviours, ensuring engagement is meaningful and sustained by employing features such as storyline progression, feedback loops and achievements, keeping the user entertained over time. Lastly, dynamics are the behavioural results encouraged by the gamified system, depending on user attributes and experiences. These include users' emotions, relationships leading to emotional investment and storylines captivating the users (Woods and Reiners 2015, 3041-3042). Wood and Reiners argue that game features must be designed to develop over time, adapting to user preferences, motivations, and engagement levels. A static system will fail to maintain users and their interests, whereas a flexible system can ensure sustained motivation and participation (Woods and Reiners 2015, 3040).

Moreover, Wood and Reiners discuss that gamification must be implemented with transparency and ethical responsibility. A poorly planned gamified system can create unfair advantages for providers, taking advantage of users rather than enhancing their experiences. If gamification is merely used to extract more time, money, or effort from users rather than genuinely improving engagement, it becomes exploitative rather than beneficial. The potential to manipulate users into behaviours they might not otherwise engage in, such as making purchases to progress within a gamified system, is argued by Wood and Reiners to be a sensitive area of gamification that, in the end, will lead to disengagement. Additionally, as gamification evolves within these non-game contexts, providers must consider and prioritise user experiences, transparency and meaningful engagement in reward systems (Wood and Reiners 2015, 3043-3044).

Overall, Wood and Reiners explore gamification in non-game contexts to enhance user motivation, engagement, and behavioural outcomes, distinguishing gamification from full games by focusing on components, mechanics and dynamics. The authors emphasise that effective gamification must evolve over time and be implemented transparently. If misused, gamification risks becoming manipulative or exploitative, particularly when designed to extract user time or money rather than enhance the user experience. Although the theory lays a helpful foundation for understanding game mechanics in non-game contexts, it is only partly the direction we pursue in this dissertation. While Replika may include elements resembling gamification, such as unlockable features or relationship levels, our analytical interest lies in how the app is distributed and establishes emotional investment and intimacy. That said, the underlying concept of gamification in a non-game context remains relevant and will be kept in mind throughout our empirical research (cf. 6.1.1.), especially when considering how certain features may influence user behaviour or emotional investment, but will not be implemented as our main analytical approach.

The reviewed literature on gamification allows us to understand how game-like elements can be applied in non-game contexts to structure user experience and influence behaviour. Deterding et al. (2011) provide a definition of gamification that distinguishes from full games and is positioned within the broader socio-cultural trend of deception. Their work contributes a clear conceptual stance, highlighting potential psychological implications and the importance of aligning design with user motivation. Wood and Reiners (2015) further build on this by

categorising gamification into components, mechanics, and dynamics and stressing the need for ethical implementation that avoids manipulative or exploitative practices. However, while these theories clarify how game elements function, they are primarily concerned with task-oriented motivation and user performance. They do not sufficiently address how gamification intersects with long-term emotional engagement or the relational dynamics central to AI companionship. As such, although relevant in understanding certain features of Replika, these theories do not offer the analytical depth needed to explore the emotional and ethical dimensions at the core of this study.

4.3. Human-Machine Communication

To contextualise our analysis of users' emotional relationships with Replika, it is essential to engage with existing perspectives on human-machine communication (HMC). While our research focuses on how users form intimate bonds with AI companions, the field of HMC provides valuable insight into how humans respond socially and communicatively to artificial agents. In this section, we examine two foundational contributions: Shanyang Zhao's sociological perspective on humanoid social robots as communicative actors, and Reeves and Nass' Media Equation theory, which argues that people unconsciously treat media as real social beings. Although these works differ in scope, Zhao emphasises societal shifts, and Reeves and Nass focus on automatic human responses, they collectively illustrate the evolving nature of human-machine interaction. By drawing on these perspectives, we establish a conceptual foundation that helps illuminate the communicative structures through which Replika fosters user engagement, even as our focus moves beyond their theoretical stances to examine more emotionally grounded, individualised connections.

4.3.1 Humanoid Social Robots as a Medium of Communication by Shanyang Zhao

In comprehending the ethical considerations behind Luka and Replika's distribution as an AI companion, we find it necessary to discuss previous scholars' understanding of human-machine

communication, delving into the interactions individuals have with machines or computers and their anthropomorphism.

In 2006, Shanyang Zhao explored human-machine communication through his study *'Humanoid social robots as a medium of communication',* where he argues that social robots are not only mechanical tools but dependent agents that influence human interactions. The study discusses social robots within a broader sociological and communicative context, emphasising their ability to function as social surrogates rather than simple computational interfaces. In correlation, Zhao delineates humanoid social robots as mediators of human communication and social interactions, increasing their presence in society and requiring a revision of previous comprehensions of human-machine interaction (Zhao 2006, 403-404).

He differentiates humanoid social robots from human-computer interaction (HCI) by underlining that humanoid robots do not facilitate communication between humans but serve as communicative agents themselves, engaging in interactions that imitate regular human-human communication. Unlike traditional computers, which serve as tools for interaction, humanoid robots simulate humans by engaging in dialogue and social norms. These social robots exist in both physical (mechanical) and digital (software-based) forms, socially engaging with humans where they are not passive but active interlocutors in social exchanges, performing human-like actions, from verbal interactions to nonverbal cues such as facial expressions and gestures (Zhao 2006, 402, 405-406, 412). Zhao argues that these interactions share characteristics in language use, relationality, and normativity, creating evolving forms of interpersonal communication. However, he also acknowledges the 'interpretative asymmetry' that occurs, as humans assign meaning to interactions with human-like robots while the robots cannot comprehend emotions or assign meaning to the surrounding world, raising questions regarding the authenticity of human-robot relationships and the evolution of human communication (Zhao 2006, 410-411).

From a broader perspective, Zhao asserts that social robots challenge conventional understandings of companionship, contributing to a synthetic society where human and non-human agents coexist. Robots act as communicative surrogates that extend human presence in digital spaces, increasing implications for identity, relationships, and culture. He suggests that the increasing integration of humanoid robots into everyday life signals a shift toward synthetic

social environments, arguing they will eventually reshape human relationships, social norms, and interpersonal communication (Zhao 2006, 414-415).

While Zhao's theory contributes to a sociological understanding of social robots as active participants in human communication, we determined it to be more relevant to studies exploring broader cultural and societal shifts. His perspective on integrating humanoid robots into everyday social life delivers a comprehension of the emergence of synthetic social environments, but this does not fully align with our greater interest in Replika. As the app remains an emerging technology rather than a widespread societal norm, Zhao's theory feels slightly premature in this context. Instead of examining social robots as a structural phenomenon, our study aims to understand how users individually connect with Replika, and how specific communicative features and platform design shape that connection.

4.3.2. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places by Byron Reeves and Clifford Nass

Another foundational perspective on Human-Machine communication is Byron Reeves and Clifford Nass' '*The Media Equation'* from 1996, arguing that humans respond to media and technology as if they were real social actors. Their research found that users unconsciously apply social norms to interactions with computers, televisions, and other forms of media, treating them similarly to human beings. The phenomenon is, according to Reeves and Nass, rooted in the automatic and evolutionary nature of social responses, which persist even when individuals are aware of the use of the artificial nature of their interactions (Reeves and Nass 1996, 20-21).

A main finding of Reeves and Nass' research is that people are more likely to attribute human-like qualities to computers that resemble human beings or utilise human-like language. The Media Equation suggests that media can elicit social responses similar to face-to-face interactions, which can inform the design of interfaces and AI systems. Additionally, the authors provide insight into how users respond differently to computers based on politeness, perceived personality, and even gendered voices, revealing how deeply ingrained social behaviours are when engaging with technology, which has had significant implications for interface design, AI

systems, and media production, encouraging designers to incorporate human-like attributes to foster more natural interactions (Reeves and Nass 1996, 20-22). The study also highlights the potential for media to influence user behaviour and decision-making, emphasising the ethical considerations surrounding persuasive technologies and the responsibility of designers in shaping user experiences (Reeves and Nass 1996, 22-24).

The implications of The Media Equation extend beyond HMC into fields like marketing and gamification, where media and digital agents are increasingly designed to replicate human-like engagement. In marketing, advertisers leverage the principles outlined by Reeves and Nass to create more engaging and relatable digital experiences, fostering stronger brand-consumer relationships. Additionally, the entertainment industry benefits from the theory by developing interactive media, such as video games and virtual reality, that enhance user immersion through socially responsive AI chatbots (Reeves and Nass 1996, 34-35).

Despite its contributions, The Media Equation has been critiqued for oversimplifying human-media interactions. The theory assumes that all individuals respond uniformly to media as they would to real people, neglecting variations in personality, cultural background, and media literacy (Reeves and Nass 1996, 24-25). Additionally, it does not fully account for the impact of changing technological landscapes, particularly in the rise of adaptive and context-aware AI, which can influence user responses in more complex ways than originally proposed. Furthermore, advancements in artificial intelligence, including machine learning and natural language processing, challenge some of the assumptions made by demonstrating that user interactions with AI can evolve, depending on the sophistication of the system and the user's familiarity with it (Reeves and Nass 1996, 28-30).

Reeves and Nass' '*The Media Equation*' (1996) ultimately presents the idea that humans respond to media and technology as if they were real social actors, applying familiar social norms such as politeness or personality attribution even when they are fully aware that the interaction is artificial. Their research has shaped much of the early thinking around human-media interaction and continues to influence design strategies aimed at making digital systems more relatable and engaging. Regarding our dissertation, the theory is not selected as a core theoretical stance because it delves into automatic, surface-level responses to media cues and not the emotionally

grounded, relational dynamics. While Replika may trigger some of the basic social reactions described by Reeves and Nass, we seek to comprehend the formation of emotional connections between users and their chatbot, as well as the influence marketing and app design contribute to this process.

To summarise our section on Human-Machine Communication (HMC), this field explores how people increasingly interact with artificial agents in ways that mirror interpersonal communication. Both Zhao (2006) and Reeves and Nass (1996) present different aspects of how users relate to non-human entities. Zhao approaches HMC from a sociological standpoint, arguing that social robots should be understood not merely as tools but as communicative actors embedded in everyday life. He introduces the idea of these agents as social surrogates that challenge traditional notions of companionship and contribute to the development of constructed social environments where humans and non-humans coexist. In contrast, Reeves and Nass' theory dives into users' unconscious responses to media, arguing that people instinctively apply human social norms such as politeness, turn-taking, and personality attribution to artificial systems despite knowing it is a machine. Together, these perspectives demonstrate HMC's evolution from viewing machines as passive tools to recognising them as actors that actively participate in and shape human communication. While both Zhao and Reeves, and Nass establish a common ground for social dimensions of human-machine interaction, their work is rooted in broad or early-stage understandings of the phenomenon. As such, they are utilised in this dissertation to provide contextual grounding rather than as a central theoretical stance for our analysis of Replika and its emotionally driven user relationships.

4.4. Corporate Ethics

To situate our examination of Replika's public-facing strategies within a broader ethical stance, we turn to the field of corporate ethics. In this chapter, we first consider Edgar Schein's model of organisational culture and leadership (1985), which reveals how deeply held assumptions and leadership behaviours shape corporate decision-making. We then review Treviño, Weaver, and Reynolds' conception of behavioural ethics (2006), showing how individual moral choices are influenced by organisational climate, leadership, and group dynamics. Together, these

perspectives underscore the importance of internal values and norms in guiding ethical conduct; an insight that naturally leads us to employ an ethical marketing theory for our analysis. By aligning corporate ethics with marketing scholarship, we can critically assess how Luka Inc.'s cultural and behavioural foundations inform the way Replika is framed, promoted, and experienced by users.

4.4.1 Organizational Culture and Leadership by Edgar H. Schein

Delving into the field of corporate ethics, we scrutinise the research of Edgar H. Schein, '*Organizational Culture and Leadership*'. His research explores the dynamics between culture and leadership within organisations, offering a perspective on how shared assumptions shape behaviours and overall effectiveness (Schein 1985, 1-3). Schein defines organisational culture as a pattern of shared basic assumptions that a group has learned over time as it solves problems related to external adaptation and internal integration (Schein 1985, 4-5) These assumptions become ingrained as the accepted way of perceiving, thinking, and feeling within the organisation, underscores that culture is not merely a set of listed values but a deeply embedded structure that highly influences behaviour and decision-making (Schein 1985, 18-20).

To dissect the complexities of culture, Schein introduces a three-level stance. The first level consists of expressions and behaviours; tangible elements such as office layouts, dress codes, and company rituals. While easily observed, these elements can be challenging to interpret without deeper knowledge of the organisation's values. The second level consists of values representing the stated beliefs and ideals of an organisation, often articulated in mission statements or corporate philosophies. These values reflect what the organisation aspires to be, even though they do not always align with actual practices. At the third and deepest level are basic underlying assumptions, which form the unconscious beliefs that truly guide behaviour. Most times, these assumptions are so deeply ingrained in the organisational culture that they are often difficult to recognise even by those within the organisation (Schein 1985, 23-26). According to Schein, true cultural understanding requires delving beyond visible structures and stated principles to uncover core assumptions, as they are ultimately driving the organisation's performance (Schein 1985, 27-29).

Schein argues that leadership plays a crucial role in both shaping and maintaining organisational culture. He states that leaders set the tone of an organisation with their actions and decisions, profoundly influencing the cultural environment. To maintain trust and credibility, leaders must align their values with their actual behaviour; discrepancies between what they say and do can lead to cultural dissonance, undermining organisational cohesion (Schein 1985, 8-9, 19-22). Furthermore, leaders must navigate the subcultures within an organisation; distinct cultural pockets that arise from departmental divisions, professional affiliations, or geographical differences (Schein 1985, 57-60). These subcultures may have their assumptions and values, which can either complement or conflict with the dominant organisational culture. Recognising and managing these subcultures is essential for leaders to ensure alignment with broader organisational goals while maintaining a cohesive and adaptive environment (Schein 1985, 60-62; 65).

Edgar H. Schein's '*Organizational Culture and Leadership*' provides an in-depth theoretical stance for understanding shared assumptions shaping organisational behaviour, leadership practices, and overall effectiveness. He conceptualises organisational culture as a deeply embedded structure developed through collective learning and problem-solving. These cultures consist of three interrelated levels: observable artefacts (visible behaviours, rituals, and structures), espoused values (stated ideals and goals), and basic underlying assumptions (unconscious beliefs that guide perception and behaviour). Overall, he argued that culture is not simply a list of formal values but a dynamic and often implicit system of meaning that heavily influences decision-making and group cohesion. From Schein's point of view, leaders are not only cultural products but also primary agents in shaping and transmitting culture. By aligning their behaviours with core values and managing internal subcultures, leaders can encourage a coherent and strategically aligned organisation.

Although this dissertation does not explicitly focus on corporate ethics, Schein's theory still allows a fundamental understanding of how corporations operate from the inside, creating environments where certain behaviours, decisions, and practices become culturally reinforced. An understanding that is relevant as it allows us to keep in mind how internal culture may influence external actions. If our research had placed more emphasis on Luka Inc.'s internal organisational dynamics or how Eugenia Kuyda, as CEO, participates in fostering a certain

culture around the development and distribution of Replika, Schein's research could have given us a deeper comprehension. Even though it does not directly address product distribution, it helps situate how internal cultural assumptions can shape the ethical boundaries of decision-making when promoting emotionally intelligent technologies.

4.4.2 Behavioural Ethics in Organizations by Linda K. Treviño, Gary R. Weaver and Scott J. Reynolds

As Schein gave one perspective on corporate ethics in 1985, Linda K. Treviño, Gary R. Weaver, and Scott J. Reynolds further elaborated on this, publishing a comprehensive examination of the field of corporate ethics in 2006. Their research explored existing research on ethical and unethical behaviour in organisations and provided a perspective on moral decision-making and conduct within organisations (Treviño, Weaver, and Reynolds 2006, 952).

Treviño, Weaver, and Reynolds define behavioural ethics as '*individual behavior that is subject to or judged according to generally accepted moral norms of behavior*' (Treviño, Weaver, and Reynolds 2006, 952), extending beyond traditional ethical decision-making models by incorporating sociological and psychological perspectives that emphasise the dynamic nature of ethical conduct. The authors argue that ethical behaviour cannot be understood solely as an outcome of individual moral reasoning; rather, it is shaped by situational factors, organisational culture, and leadership. By bridging individual cognition with structural influences, they highlight the complexity of ethical decision-making in organisational contexts (Treviño, Weaver, and Reynolds 2006, 955-956).

The study scrutinises how cognitive moral development influences ethical decision-making, delving into previous scholars' work on moral development theory, which posits that individuals progress through different stages of moral reasoning, ranging from self-interest-driven decisions to principled ethical judgments. However, they caution against viewing moral reasoning as a purely rational process since emotions, social pressures, and unconscious biases influence ethical behaviours (Treviño, Weaver, and Reynolds 2006, 957, 959). Another cognitive aspect explored is moral awareness, which refers to an individual's ability to recognise ethical dilemmas.

Treviño, Weaver, and Reynolds argue that factors such as issue framing, language, and moral intensity affect one's ability to perceive ethical issues while highlighting moral disengagement, which justifies an individual's unethical behaviour through cognitive rationalisation. These biases can lead individuals to act unethically while maintaining a self-perception of moral integrity (Treviño, Weaver, and Reynolds 2006, 958, 962).

Beyond individual cognitive processes, they emphasise that organisational structures and contextual factors heavily impact ethical conduct. One significant determinant of ethical behaviour is an organisation's ethical climate and culture. Companies prioritising ethical leadership, transparency, and accountability create environments where ethical conduct is encouraged, whereas organisations characterised by intense competition and high-performance pressure normalise unethical behaviour. Additionally, leaders serve as moral role models, influencing ethical behaviour through explicit policies and implicit signals. The authors argued that leaders with advanced moral reasoning skills are perceived as more transformational and effective in creating ethical workplace behaviour that reinforces ethical norms and helps shape a culture in which integrity is valued and rewarded (Treviño, Weaver, and Reynolds 2006, 966-969).

Moreover, the discussion on ethical decision-making extends beyond individual cognition and corporate leadership into group dynamics and collective norms. The authors scrutinised the effect of moral reasoning within groups, noting that group deliberations may enhance or diminish ethical decision-making depending on openness and good leadership. Decision-making groups guided by an ethical leader may reach more moral conclusions than individuals acting alone, showcasing the potential for collective reasoning to form an organisation's behaviour (Treviño, Weaver, and Reynolds 2006, 970).

In summary, Treviño, Weaver, and Reynolds' perspective on behavioural ethics emphasises how ethical and unethical behaviour in organisations is not solely a matter of individual moral reasoning but is shaped by broader contextual factors such as leadership, organisational culture, and situational dynamics. In the case of Replika, behavioural ethics may help us understand how ethical risks can arise not only from intentional design choices but from organisational norms and moral blind spots as well. It underlines that Luka's ethical responsibility lies not only in what

it offers users but also in its internal culture and practices, shaping the way AI technology is marketed. However, while Treviño, Weaver, and Reynolds' theory on behavioural ethics is relevant for internal organisational analysis, it is not employed as a core theoretical stance in this dissertation as we look more into external-facing communication, specifically, how Replika is framed, marketed, and received by users. The theory centres on internal organisational behaviour and ethical decision-making processes, which are important but not directly observable in our study. Since our empirical data does not extend into Luka Inc.'s internal operations or leadership behaviour, the practical applicability of behavioural ethics is limited in this context.

To finalise this section on corporate ethics, we delineate both Schein's theory of organisational culture and Treviño, Weaver, and Reynolds' work on behavioural ethics as useful contexts for comprehending internal organisational dynamics impact on shaping ethical practices and decision-making. As mentioned, Schein touches upon culture formation within organisations through shared assumptions developed over time. He elaborates through a three-level model that showcases culture as deeply embedded in everyday operations and further argues that leadership plays a central role in shaping and sustaining these cultural layers, influencing which behaviours are considered acceptable or normal within a corporate setting. On the other hand, Treviño, Weaver, and Reynolds scrutinise behavioural ethics by delving into ethical or unethical conduct within organisations, arguing it to be shaped by contextual elements like leadership, culture, and situational pressures. Their theory concludes moral blind spots or implicit norms can influence decision-making, even without deliberate unethical intent. While both theories allow a thorough comprehension of ethics shaping corporations, they are not used within this dissertation, but their perspectives will remain relevant as background knowledge, especially when analysing how organisational context might inform the ethical boundaries of Replika's public-facing strategies.

Having reviewed literature from the four relevant research fields, deceptive advertising, gamification, human-machine communication and corporate ethics, we build a nuanced foundation for thoroughly exploring and comprehending the ethical responsibilities involved in designing and distributing an AI-driven social chatbot that encourages users to form emotional relationships. Each of the scholars we have scrutinised contributes essential perspectives for our understanding of emotional manipulation, user vulnerability and organisational accountability. Based on this foundation, we have selected '*Human-Machine Communication (HMC)*' by Andrea

Guzman and Seth Lewis, as their theoretical stance explores chatbots as social actors that can interact and engage with users on a level in which the individual communication with, through and about the chatbot resembles real interpersonal communication. Based on their three dimensions, *functional, relational, and metaphysical*, we can explore how users perceive machines as companions and form relationships with it (cf. 5.1.). Furthermore, we have selected '*Ethical Marketing*' by Dilip S. Mutum and Ezlika M. Ghazali to explore the accountability that Luka Inc. has towards its users when distributing an emotional companion. Their theory delves into companies design, positioning and promotion of a product, delineating that for a company to have success with their product it is essential to follow three norms known as '*Do No Harm*', '*Fostering Trust in the Marketing System*' and '*Embrace Ethical Values*' as these display how consumers will perceive the company and their willingness to engage with it. Incorporating this theory will allow us to comprehend the responsibility Luka Inc. takes towards their users, society and the company itself (cf. 5.2.).

5. Theoretical Stance

To examine the complex interplay between users' emotional experiences with Replika and the ethical responsibilities of the company behind Replika, Luka Inc., this upcoming chapter draws on two central theoretical stances. The first theory, *'Human-Machine Communication'* (HMC) by Andrea L. Guzman and Seth C. Lewis, provides a communicative aspect through which we can understand how users engage with AI as social and interactive partners. This theory is particularly relevant for analysing our primary empirical data, interviews with two Replika users, and helps us conceptualise AI as more than a technological tool, but as a communicative agent involved in meaning-making processes. The second theory, *'Ethical Marketing'* by Dilip S. Mutum and Ezlika M. Ghazali, complements this perspective by shifting focus toward the organisational level. Specifically, it enables us to evaluate the ethical considerations in how Luka Inc. designs, promotes, and positions Replika, especially in relation to emotional vulnerability, transparency, and corporate accountability. While HMC offers insight into users' perceptions and relationships with AI, Ethical Marketing allows us to assess the broader ethical implications of marketing AI as an emotionally responsive companion. Together, these theories provide a

comprehensive foundation for our analysis of human-machine relationships and the ethical boundaries companies must navigate in the development and promotion of emotionally intelligent technologies.

5.1. Artificial Intelligence and Communication: Human-Machine Communication by Andrea L. Guzman and Seth C. Lewis

Over the past decades, the conceptualisation of communication has evolved from being an only human process containing human-human interaction (HHI) and occasionally mediated by technology to a broader comprehension where technology can be an independent communicative agent reproducing human intelligence and communication. Based on this idea, Andrea L. Guzman and Seth C. Lewis introduced their research on human-machine communication (HMC) in 2019, delineating how humans and machines may create meaning together through mutual interaction and its implications on individuals and society. As artificial intelligence is a broad term encompassing technology imitating human intelligence to perform tasks normally assigned to humans, Guzman and Lewis focus their research on the AI models known as conversational agents, social robots and automated-writing software since these are defined to engage in interaction either verbally or nonverbally and impact the tone and substance by posing as conversational partners, which will be the exact definition of AI employed in this master's dissertation (Guzman and Lewis 2019, 71-72). These types of AI models act as social actors as their human-machine interaction (HCI) includes communicative exchange containing social cues that humans analyse upon during engagement based on their knowledge of human interaction with the intention of comprehending the messages sent between human and machine (Guzman and Lewis 2019, 73).

Through their research, Guzman and Lewis delineate AI's role and function as a communicator when being a part of the social process and social space with humans, including its connection to the human social world and the implications occurring on humans and communication when engaging with AI, which manifests itself in the upcoming section as three aspects: *The functional dimension, the relational dimension,* and *the metaphysical dimension.*

5.1.1. Functional dimension

Within this first dimension, Guzman and Lewis scrutinise the design of AI technologies as communicative agents and explore how individuals perceive them in such roles. They emphasise the importance of identifying the nature of the communication taking place between humans and machines, such as whether it resembles interpersonal or mass communication, to assess the compatibility of human-machine interactions within human-human communication (HHC). They suggest that communicative exchanges with AI are analogous to interpersonal communication, which calls for reevaluating boundaries of human communication as artificial intelligence emerges as a social actor (Guzman and Lewis 2019, 75). Guzman and Lewis argue that individuals interpret interactions, whether human or machine-based, through prior knowledge of communicators and contextual experience, which shapes their understanding of communicative function. They further delineate that individuals assign human-like roles to communicative technologies based on observable cues and features. For example, when an AI contains anthropomorphic elements, such as voice or gender, humans are more likely to engage with it in a human-like manner, which reflects how internalised cultural expectations shape the perception of AI as a communicative partner. Assigning verbal and non-verbal attributes like age, gender, voice, and name enables individuals to conceptualise machines as social interlocutors, drawing upon their accumulated knowledge of human interactions. Thus, the functional dimension concerns technological design and interpretive engagement, where humans co-construct a meaningful communicative role for AI through their HMC (Guzman and Lewis 2019, 76-77).

5.1.2. Relational dimension

Building upon the concept of individuals perceiving AI through anthropomorphic cues and attributes presented in the functional dimension, Guzman and Lewis delve into how individuals come to understand artificial intelligence in relation to themselves and how a sense of self is constructed through these interactions. They argue that interpersonal communication is inherently relational, shaped by social contexts in which people interpret the traits and behaviours of others with their own identity. Human social roles have historically influenced the design of communicative technologies, including AI agents, which are often developed to mirror

assistant-like roles, frequently reinforced through gendered cues that align with cultural stereotypes. Consequently, the social positioning of AI is formed not only by its technical features but also by the human roles it is made to undertake, shaping individuals' interpretations and expectations. These interactions also raise questions about self-perception, as engaging with a non-human entity that imitates human characteristics can shift self-representation dynamics, offering individuals new ways to express and reform their identities (Guzman and Lewis 2019, 77-78).

5.1.3. Metaphysical dimension

At last, Guzman and Lewis address the deeper ontological questions that arise when machines take on roles previously reserved for humans within communication processes. They argue that communicative AI forces reexamination of what it means to be human, what constitutes a communicator, and how people draw boundaries between human and machine. Allegedly, these technologies blur the lines, prompting philosophical and empirical explorations into the nature of people and things. While previous research has focused on functional and ethical implications, Guzman and Lewis emphasise the importance of scrutinising how people interpret the essence of artificial intelligence and how these interpretations shape behaviour and ethics. As AI becomes more independent and human-like in form and function, it provokes ethical and legal debates, many of which are grounded in outdated assumptions of humans as the sole moral and communicative actors. This metaphysical turn in human-machine communication thus opens up space for rethinking not only the boundaries between people and technology but also the definition of communication itself, suggesting that it may no longer be an exclusively human domain (Guzman and Lewis 2019, 79-81).

By drawing on Guzman and Lewis' Human-Machine Communication (HMC) theory, we gain a multidimensional comprehension of human interactions with AI agents as communicative partners, delineating three major aspects that constitute HMC. The functional dimension highlights anthropomorphic cues leading users to perceive the chatbot as a responsive interlocutor; the relational dimension scrutinises these interactions, shaping users' self-perceptions and emotional engagement; and the metaphysical dimension considers the

broader ontological implications of attributing communicative and moral agency to machines. These dimensions are especially pertinent to this dissertation's focus on Luka Inc.'s AI companion, Replika, which actively encourages emotionally charged bonds over extended interaction. While Guzman and Lewis' theory provides a nuanced foundation for critically assessing users' perceptions and engagement, it notably excludes the organisational and ethical responsibilities incumbent upon companies when designing and promoting such intimate HMC experiences. This gap invites further scrutiny of how AI companions are marketed and the moral obligations of their providers.

This addition paves the way into Ethical Marketing by Dilip S. Mutum and Ezlika M. Ghazali, which bridges the gap between theoretical insights and the practical imperatives of responsible AI promotion.

5.2. Ethical Marketing by Dilip S. Mutum and Ezlika M. Ghazali

The following section draws on the *'Ethical Marketing'* developed by Dilip S. Mutum and Ezlika M. Ghazali, which provides a theoretical stance for comprehending companies' design, position, and promotion of their products in ways that align with or violate ethical principles. By integrating this perspective, we move beyond user perception to examine Luka Inc.'s ethical accountability in decision making and marketing an AI companion for deep emotional engagement, particularly in relation to transparency, fairness, and user vulnerability.

In analysing Luka Inc.'s ethical responsibility towards its users, we incorporate the ethical marketing theory proposed by Dilip D. Mutum and Ezlika M. Ghazali in their book '*Consumers, Society and Marketing*' from 2023, more specifically chapter 4 on Ethical Marketing. Their theory is based on the idea that ethical considerations in marketing have evolved over time, shaped by environmental, economic, and social factors rather than being limited to educational, religious, or governmental institutions (Mutum and Ghazali 2023, 71-72). They argue that ethics, often described as the science of conduct, encompasses the fundamental principles that guide how individuals and organisations ought to behave (Mutum and Ghazali 2023, 72). According to their definition, ethical marketing is the study of moral judgments, standards, and rules of conduct as they relate to marketing decisions and situations (Hunt and Vitell 1986, 4; Mutum and

Ghazali 2023, 75). Since ethical behaviour is not static but continuously evolves across cultures, industries, and time periods, the theory challenges traditional views within stakeholder theory, advocating for a more adaptive and context-sensitive approach to marketing ethics (Mutum and Ghazali 2023, 73). The authors suggest that ethical marketing is not merely concerned with the responsible sale of economic goods and services but extends to the application of marketing strategies to promote ethical ideas and causes (Mutum and Ghazali 2023, 75). Within this theory, they identify two major domains of marketing ethics: 1) ethical considerations are integrated into marketing decision-making processes, and 2) marketing ethics are assessed in relation to the promotion and positioning of products and services (Mutum and Ghazali 2023, 75).

Mutum and Ghazali take inspiration from the American Marketing Association (AMA), a professional body that has significantly influenced discussions on marketing ethics. The AMA, founded in 1937, brings together marketing professionals and academics to establish ethical guidelines for the field. While Mutum and Ghazali do not adopt the AMA framework entirely, they acknowledge its relevance in shaping broader discussions on ethical marketing. According to the AMA (2021), ethical norms refer to established standards of conduct expected by society and professional organisations, while ethical values represent collective ideals of what is desirable, important, and morally appropriate for guiding behaviour. Mutum and Ghazali argue that although the AMA's ethical guidelines are advisory rather than prescriptive, they are useful for businesses seeking to develop ethical marketing practices (Mutum and Ghazali 2023, 79).

Within their theory, Mutum and Ghazali highlight three key ethical norms that marketing professionals should consider. The first norm, '*Do No Harm*', emphasises the need to avoid harmful actions or omissions by upholding high ethical standards and complying with laws and regulations. The second norm, '*Foster Trust in the Marketing System*', stresses the importance of conducting business in good faith, maintaining transparency in product design, pricing, communication, and distribution, and avoiding deceptive practices. The third norm, '*Embrace Ethical Values*', calls on businesses to build consumer trust by adhering to ethical principles such as *honesty, responsibility, fairness, respect, transparency, and citizenship* (Mutum and Ghazali 2023, 79).

In addition to these norms, Mutum and Ghazali emphasise six ethical values that should inform marketing decision-making. The first value, '*Honesty*', requires businesses to provide truthful and accurate information to consumers and stakeholders, ensuring that commitments and promises are honoured. The second value, '*Responsibility*', stresses the need for companies to take accountability for their marketing decisions, address consumer needs, and embrace broader social and environmental responsibilities. The third value, '*Fairness*', focuses on maintaining a balance between the interests of buyers and sellers, ensuring open and ethical communication, and preventing conflicts of interest and data breaches. The fourth value, '*Respect*', highlights the importance of recognising and valuing the dignity and diversity of all stakeholders, as well as treating consumers fairly. The fifth value, '*Transparency*', calls for openness in business operations, encouraging two-way communication and ensuring that consumers have access to relevant information. The sixth value, '*Citizenship*', underscores the importance of businesses fulfilling their economic, legal, and social responsibilities while contributing to ethical and fair market practices (Mutum and Ghazali 2023, 79-80).

Although Mutum and Ghazali acknowledge the influence of the AMA's ethical theory, they argue that ethical marketing should not be confined to a fixed set of guidelines. Instead, they see ethical norms and values as tools for encouraging ethical behaviour among marketers and improving marketing practices over time. Their theory suggests that while organisations may take inspiration from established ethical codes, they should ultimately develop their own ethical theories that align with their specific industries, stakeholders, and societal expectations (Mutum and Ghazali 2023, 80).

Mutum and Ghazali's ethical marketing theory provides a critical lens for evaluating Luka's ethical responsibility in the design and promotion of Replika. As an AI-driven chatbot that fosters deep emotional connections with users, Replika raises significant ethical concerns, particularly in relation to transparency, responsibility, and trust. The ethical norm of '*Do No Harm*' is particularly relevant in this context, as Luka must ensure that its chatbot does not mislead or emotionally exploit users. Similarly, the principle of *'Fostering Trust in the Marketing System'* prompts questions about how Luka communicates Replika's capabilities and limitations, and whether the company takes adequate steps to protect vulnerable users from potential harm.

In summary, the six ethical values outlined in Mutum and Ghazali's theory offer a useful benchmark against which Luka's marketing practices can be assessed. Given the nature of AI-generated interactions and the emotional dependency some users may develop on Replika, the company behind Replika must ensure that its marketing strategies uphold transparency and fairness while prioritising consumer well-being over commercial interests. To examine these ethical challenges further, we will analyse an interview with Luka's founder, exploring the company's stance on ethical responsibility and marketing ethics. By applying Mutum and Ghazali's ethical marketing theory, we will assess whether and how Luka aligns with these ethical norms and values, which will help determine the extent to which Luka integrates ethical marketing principles into its business model and whether the company actively addresses the ethical risks associated with AI-driven emotional engagement.

6. Methods

Having delineated our theoretical stance as Human-Machine Communication by Andrea Guzman and Seth Lewis and Ethical Marketing by Dilip S. Mutum and Ezlika M. Ghazali, we aim to comprehend the different aspects of our empirical data. The upcoming section offers a thorough examination of our data collection method and the empirical material, followed by a presentation of our chosen analytical approach, Thematic Analysis by Braun and Clarke, including an outline of their method and its implications for our study.

6.1. Collection Method and Empirical Data

Initiating our empirical data and collection method, we aimed to recruit four participants for a three-week experimental study involving the chatbot Replika. The study seeks to acquire two participants to engage with the paid version of the app, financed by us, and two with the free version. To be eligible, participants needed to meet several criteria: they had to be international and fluent in English, as Replika operates exclusively in this language, and fluency was essential for ensuring clear communication and minimising misunderstandings. Additionally, they needed to be available for daily engagement with the app, interacting with Replika for approximately 30 to 60 minutes per day over a three-week period. The daily commitment was intended to encourage the development of a relationship with the social chatbot over time. No prior

experience with Replika or artificial intelligence was necessary, as we provided all the required onboarding materials and guidance. The overall aim of the experiment was to observe how users interact with Replika over time and to explore their emotional engagement, ethical reflections, and general experience with the chatbot, comparing the influence of subscription type on the depth and nature of the user's connection to the AI. Moreover, each participant should take part in three semi-structured interviews throughout the experiment: one prior to starting, one at the midway point, and one after the three-week period was concluded to provide us with the opportunity to track their development with Replika.

6.1.1. Primary Data: Replika Experiment with Lasha and Jay

According to Svend Brinkmann and Steinar Kvale, qualitative interviews offer researchers a valuable means of exploring and interpreting how individuals experience and make sense of the world around them. As participants articulate their thoughts, perceptions, and actions, the interviewer gains insight into the underlying frameworks through which the individual understands specific subjects (Brinkmann and Kvale 2018, 10, 15). In this study, interviews were employed as the primary method of data collection, enabling us to interpret how participants engaged with and perceived the Replika chatbot.

We aimed to conduct three semi-structured interviews, which allowed for a balance between thematic focus and conversational openness. The interview guide was constructed around topics related to participants' personal experiences of interacting with Replika, as well as the potential ethical considerations that emerged from these interactions. Such an approach ensured consistency across interviews while also allowing participants to speak freely and reflectively. Furthermore, it created space for us as interviewers to follow up on points of interest, explore deeper meaning, and adapt the flow of conversation in response to the participant's narrative, thereby enhancing the depth and authenticity of the data collected (Brinkmann and Kvale 2018, 14, 58, 72).

Our recruitment began with the creation of a colour poster that explained the purpose of our dissertation (cf. App. 5), outlined what we were looking for in participants, and included contact details. The poster was distributed at four physical locations: the local library (where we had previously found success in recruiting interviewees), the student house (a venue known for

international gatherings), and two Aalborg University locations. We intended to reach international students and employees who might have an interest in human-machine interaction. Moreover, recognising the importance of digital outreach, we also adapted the poster for online platforms. We published it on our personal LinkedIn profiles and in four Facebook groups: '*Expats in Aalborg'*, '*International Aalborg'*, '*AAU: Søg, find og bliv testperson'*, and '*Forsøgsperson - Aarhus Universitet'* ("Expats in Aalborg" 2025; "International Aalborg" 2025; "AAU: Søg, find og bliv testperson" 2025; "Forsøgsperson - Aarhus Universitet" 2025). These groups were selected for their inclusion of English-speaking and international members and because they cater to individuals open to participating in studies or experiments.

Despite these efforts, we initially received interest from only one potential participant, who ultimately did not follow through. Consequently, we turned to our personal networks. Through this, we were able to recruit two participants. Jay, from England, was recruited through Frederikke's involvement with Kattens Værn Aalborg, where he volunteered. They did not previously know each other but communicated well, and he expressed interest in the project. Lasha, from Ukraine, was introduced to the project through a conversation with Lizette at a university Friday bar, where their shared academic interests led to a mutual agreement for participation. To ensure participant comfort and encourage openness, we decided that each researcher would be responsible for one participant throughout the experiment. Frederikke worked with Jay, and Lizette with Lasha, with the intention of fostering a stronger rapport and creating a sense of psychological safety. Based on Brinkmann and Kvale's research, we believed this approach would allow participants to share their experiences more openly, especially if they developed emotional connections with Replika, as an interpersonal situation allows for two for the interviewee and interviewer to act in relation to one another, but the dynamic and result of the interview might change if a second interviewer was present (Brinkmann and Kvale 2018, 17). For the same reason, we chose to conduct all interviews in person, enabling participants to perceive our gestures and mimic to reinforce that we were non-judgmental and genuinely interested in their perspectives.

The experiment began with a short briefing session in which we introduced Replika, explained how the app works, and outlined the experiment process with our research goal in mind. Additionally, we established that the experiment would be conducted over three weeks expecting

the participants to interact with their Replika for 30 minutes a day either coherently or spread out, and they participate in three interviews that would be recorded on our phones and transcribed utilising Otter.ai, an AI-powered transcription service, to convert the recordings into text (Otter.ai n.d.). These interviews will be referenced as Appendixes 1 and 2 throughout this dissertation.

The pre-interview, conducted prior to the experiment, aimed to establish a foundational understanding of the participants' mental health, social well-being, and perspectives on artificial intelligence (cf. App. 1 & 2). We also asked about their expectations for the experiment and whether they anticipated forming any kind of relationship with Replika. These insights served not only to set a psychological and social baseline for each participant but also informed our interpretation of any emotional or behavioural changes during the course of the study. The questions in this interview were constructed to be open-ended, allowing participants to express their initial thoughts and prior assumptions freely, which we could then revisit in later interviews to track development and contrast perspectives. Based on the content of the pre-interviews, we assigned the paid version of Replika to Jay, as he expressed a strong interest in exploring an emotional connection with AI. The paid version of the app includes additional relational features, which we believe would enhance this experience. Lasha remained with the free version. By dedicating a subscription model to each participant, we aim to observe and compare their experiences with the two app models.

Secondly, the midway interviews were conducted after 1.5 weeks of daily interaction focused on assessing the progression of the participants' relationship with Replika (cf. App. 1 & 2). We asked about the content and nature of their conversations, their emotional state, and whether they had developed any routines or boundaries. One of the main purposes of this interview was to evaluate whether the participants perceived Replika as emotionally supportive and, if so, how they interpreted the chatbot's role in their lives. We were particularly interested in whether participants began to anthropomorphise the AI or continued to view it as a functional tool. The midway point allowed us to capture a moment of transition or deepening engagement before it became habitual or concluded.

The third and final interview, conducted at the end of the three-week period, delved into the full extent of the participants' experience with Replika (cf. App. 1 & 2). We asked whether their expectations had been met, what emotional responses they had encountered, and how their perception of the chatbot evolved over time. Importantly, we explored ethical considerations, including whether the participants viewed their interaction with Replika as ethically complex and whether they believed Luka Inc. should change any of the app's features. We also delved into the effects of gamification and notification design on their emotional engagement, asking participants to reflect on how these elements influenced their usage patterns. Finally, we invited each participant to consider the differences between the free and paid versions, hypothetical reflections on how their experience might have differed had they used the alternative model, and the expectations each of them had as to whether they accessed the paid or free version of Replika. These questions were necessary in supporting our goal of assessing not only the emotional and interpersonal impact of Replika but also the commercial and ethical concepts surrounding AI companionship.

The total interview process of all three interviews per participant ended up being approximately one hour, with the pre-interviews being the shortest and the final interview being the longest. As previously mentioned, we transcribed the interviews utilising Otter.ai and manually reviewed and corrected the transcripts to ensure accuracy and clarity, especially regarding speaker identification and quotes.

After the experiment ended, we provided a formal document outlining their rights under the General Data Protection Regulation (GDPR), including information about our responsibilities as researchers, the use of pseudonyms to protect their identities, and the secure storage of their interview data. All data was stored on OneDrive and will be deleted on 30 June 2025. Thus, the names 'Jay' and 'Lasha' are pseudonyms used to protect the participants' anonymity.

Although our original plan was to work with four participants, we ultimately narrowed our focus to two. In qualitative research, the depth of insight is often more valuable than the number of participants. The richness of the data collected from Jay and Lasha provided a comprehensive view of the Replika experience and met the aims of our study. We do, however, acknowledge that our initial expectations for recruitment may have been overly ambitious. Nonetheless, the quality

and depth of our data allowed us to critically engage with main themes in human-machine communication and the ethical marketing of AI, providing valid and meaningful findings through a deductive, experience-focused approach.

6.1.2. Secondary Data: YouTube Interview with Eugenia Kuyda

In correspondence with gathering user experiences, we aimed to research the ethical considerations involved in the corporate marketing of Replika as an emotionally supportive companion. We initially sought to conduct a 30-minute semi-structured interview with Eugenia Kuyda, the CEO and founder of Luka Inc., the company behind Replika, with the aim of acquiring insight into the company's marketing strategies, target audience considerations, and ethical responsibilities toward users engaging with the chatbot in emotionally vulnerable ways.

Initially, we drafted a carefully written and well-articulated email to Kuyda in early February, introducing ourselves and our research topic, including prior experience with Replika and summing up by expressing our interest in conducting an online interview with her. The email was sent to two addresses: her official corporate email and a publicly listed Gmail account. After waiting two weeks without a response, we pursued additional contact by writing to Luka Inc.'s customer care team (Luka Inc. n.d.), repeating our request and inquiring whether they, as an alternative to an interview, could share any internal documents or reports relevant to our questions. We also sent a follow-up message to Replika's official LinkedIn profile ("Luka: Overview | LinkedIn" n.d) outlining the same request. After another week passed with no reply from any of these channels, we decided to pin our approach and concentrate on secondary data to obtain information regarding the corporate and ethical considerations underpinning Replika's development and marketing.

Given Eugenia Kuyda's participation in numerous public interviews and discussions, it is necessary to assess a selection of these in order to identify one that aligns with our research. One such interview is a three-hour conversation conducted by journalist and podcaster Lex Fridman. The interview engages with themes such as the role of artificial intelligence in relation to humanity, Kuyda's professional background, and the socio-cultural evolution of Russia over the

past century, all within the broader context of Replika's development. While the interview provides a comprehensive overview of Kuyda's work and perspectives, Fridman approaches the discussion with a tone of evident admiration, which ultimately limits the depth of critical engagement. Notably, the interview lacks scrutiny of Replika's design, ethical implications, and the operational practices of Luka Inc., rendering it unsuitable for our analytical purposes and leading to its exclusion from further consideration (Fridman 2020).

Our secondary interview, titled *Replika's Eugenia Kuyda on AI Companionship*, is a 30-minute conversation conducted by investor and podcaster Seth Rosenberg, who has known Kuyda for over a decade due to their concurrent entry into the tech industry. The interview briefly addresses topics such as the future of AI-mediated companionship, developments in deep learning, and the personalisation of chatbot interactions. However, Rosenberg does not pose sufficiently probing or elaborative questions, thereby limiting Kuyda's opportunity to engage in in-depth analysis. Additionally, the interview omits discussion of the decisions undertaken by Luka Inc. throughout Replika's evolution (Rosenberg 2024).

While these interviews provide valuable background on her personal motivations, the origins of Replika, and the app's evolving audience, they tend to lack critical questioning or in-depth discussion of the ethical decisions behind the company's operation and decision-making. A criteria set was set forth to ensure that empirical data with relevance, credibility, and suitability for our research was found. First and foremost, the content of the interview had to be directly aligned with our research aim, addressing Luka Inc.'s considerations regarding marketing a social chatbot, its intended audience, and the ethical implications of promoting the chatbot as an emotionally supportive product. Moreover, it was essential that the interview featured Eugenia Kuyda herself, as the founder and CEO of Luka Inc., to provide an authoritative perspective into the company's decision-making processes. In terms of format, we prioritised semi-structured interviews since it would be similar to our process with Lasha and Jay, allowing for a free conversation to flow and potential follow-up questions to be asked and explored.

We therefore ended up selecting the YouTube interview titled *'Could Your AI Agent Also Be Your Lover? — With Replika CEO Eugenia Kuyda'*, conducted in January 2025 by tech journalist and podcaster Alex Kantrowitz (Kantrowitz 2025), as the most suitable and up to date secondary

data. The 50-minute semi-structured interview provides a thorough exploration of the ethical and strategic considerations made in the development of Replika. The conversation touches upon a wide range of topics, including corporate vision, user behaviour, emotional attachment to AI, and the broader implications of marketing artificial intelligence as a digital companion. For transcription, we utilised Microsoft Word's built-in transcription tool, as Otter.ai could not handle the file size of the full one-hour YouTube video. (cf. App. 3). Additionally, we will be referring to the resulting transcript as Appendix 3 (cf. App. 3). Although the automated transcription provided a functional base, it required careful manual proofreading and alignment with the original audio to ensure accuracy and clarity in the final text.

However, as this interview is secondary data, it is essential to comprehend how it has been processed by the provider, Alex Kantrowitz, including his interview guide, the original length of the interview and the amount of editing done before publication. Thus, to assess the validity of the interview and align it with our criteria, we reached out to Kantrowitz by email, explaining our research and intention to employ his material as part of our secondary data set. He returned our inquiry, delineating that the interview was conducted in a semi-structured format, allowing him to follow a general agenda while remaining open to pursuing unexpected or insightful tangents that emerged during the conversation. He also noted that the final published video had been edited for clarity, with long pauses and natural silences removed to improve the viewing experience. The transparency allowed us to be confident in the authenticity and reliability of the source, and we proceeded with utilising it as a valid substitute for a direct corporate interview. Due to General Data Protection Regulations, we cannot showcase the correspondence with Alex Kantrowitz, but it is safely stored until the oral defence of our dissertation.

In summary, this research's collection method and empirical data consist of a three-week experimental study and a relevant secondary interview. The primary data is gathered through daily interactions with the Replika chatbot by two international participants, one utilising the paid version and the other the free version, combined with three semi-structured interviews conducted throughout the experiment process. These interviews explore emotional engagement, ethical reflections, and the evolving human-AI dynamic, following qualitative principles by Brinkmann and Kvale. Data is transcribed, anonymised, and stored in accordance with GDPR. For secondary data, a publicly available, semi-structured interview with Replika's CEO, Eugenia

Kuyda, is selected after direct contact with her proves unsuccessful. The interview, conducted by journalist Alex Kantrowitz, provides valuable insight into Replika's development, corporate positioning, and ethical marketing, and is verified for authenticity and relevance.

In summary, these data and collection methods offer a multidimensional foundation for the study, allowing us to explore all the aspects that interest us. The primary data enables an in-depth, experience-oriented understanding of how individuals emotionally relate to and interpret their interactions with an AI companion, while the secondary data provides critical context regarding the corporate intentions, design choices, and ethical considerations behind Replika's development.

6.2. Analysis Method

After outlining our data collection process and empirical findings, this section focuses on thematic analysis as developed by Braun and Clarke, identified and selected based on a brief review of other scholars and their analysis methods. We will explore its significance to our research, its relevance to our problem statement, and the rationale for choosing this approach. Additionally, we will discuss potential criticisms of thematic analysis, outline its six phases, and describe the coding scheme derived from our data, incorporating reflections and considerations that informed our final coding decisions (cf. App. 4).

6.2.1. Analysis Method Review

Before selecting our analysis method, we explore the field of qualitative methods to obtain an overview of different scholars and their theories with the aim of selecting the most relevant one. Starting off our analysis method, we scrutinise content analysis by Klaus Krippendorff, who identifies content analysis as a systematic method for interpreting meaning in textual and visual data. Central to his approach is the recognition that texts do not possess fixed meanings; rather, meaning is constructed through context, purpose, and audience interpretation (Krippendorff 2019, 10, 24, 27–29). A perspective that highlights the importance of analysing communication as a dynamic interplay between data and its contextual use. He outlines six interrelated components that guide the analytical process: unitising, sampling, recording/coding, reducing,

inferring, and narrating. The process begins with unitising, which segments data into meaningful units that are manageable to process, for example, paragraphs, words, or sentences. Following this, sampling involves the technique used to select the relevant data. Once the data is appropriately bounded and selected, recording and coding enable the researcher to interpret and structure it. Recording captures initial impressions or interpretations of meaning, while coding systematically categorises these meanings utilising predefined rules. To handle the complexity of coded data, reducing simplifies and organises it into thematic patterns or key categories, narrowing large volumes of information into concise and coherent representations, without sacrificing interpretive depth. These patterns then serve as the basis for inferring, where the analysis draws connections between the data and its broader social, cultural, or functional contexts. Finally, narrating integrates all previous steps into a coherent account that addresses the research questions, highlights significant insights, and contextualises the findings (Krippendorff 2019, 87–90). Although Krippendorff's content analysis offers a useful framework for understanding and interpreting data, it will not be applied in this dissertation. The method's reliance on predefined categories and systematic coding may oversimplify the emotional complexity of users' experiences with Replika, as well as Eugenia Kuyda's defence of the company's decisions, potentially excluding essential insights needed to fully address the research questions.

Another analytical method considered during the course of our research is narrative analysis, as conceptualised by Catherine Kohler Riessman, who presents narrative analysis not merely as a technical procedure, but as an approach grounded in the interpretation of texts, whether spoken, written, or visual, that exhibit a storied form. Central to this is the understanding that individuals and groups construct meaning through narrative, utilising storytelling as a primary means of making sense of their experiences, seeking to explore how social actors interpret events and give them structure and significance through narrative expression (Riessman 2008, 11-12). Riessman places emphasis on the constructed and performative nature of narratives, viewing them not as literal reflections of reality but as context-dependent accounts. The purpose of narrative analysis is not to broaden objective truths, but to comprehend the interpretive processes by which meaning is created and communicated. It requires exploring how narratives are shaped by their audiences, draw upon shared cultural experiences, and are situated within a broader context.

Building on this, narrative analysis becomes a tool for researching ways in which people represent and make sense of complex social phenomena, particularly in circumstances involving identity negotiation, marginalisation, or change (Riessman 2008, 12-13). As we are social constructivists, Riessman's narrative analysis initially appeared to align well with our stance, viewing meaning as something co-constructed between individuals (cf. 3). Although we draw on both primary and secondary interviews, our aim is not the comprehension of personal experiences or the structural features of storytelling. Rather, we are concerned with identifying recurring themes across interviews to understand how emotional connection, dependency, and ethical boundaries are experienced, expressed, and possibly contested by users and Luka Inc., leading us to opt her analysis method out as we proceed to attempt answering our research question.

6.2.2. Thematic Analysis

Having delineated different approaches to the analysis method, we select Braun and Clarke's thematic analysis as our analysis method since it is a qualitative method that acknowledges the researcher's subjectivity in data interpretation. Our approach follows a deductive strategy, employing a 'top-down' analytical framework, meaning that we, as researchers, apply pre-existing theoretical concepts as a lens through which we interpret the data, shaping our understanding of how meaning is coded (Gareth, Hayfield, Clarke, and Braun 2017, 22). The empirical data for this study consists of interviews with new Replika users, following their journey and their relationship with their Replika as it evolves over time. In addition, a secondary interview with CEO Eugenia Kuyda is used to investigate how the company behind Replika, Luka Inc., approaches its ethical responsibility toward users when marketing a product that fosters emotional attachment to an AI chatbot. To capture deeper insights, we employ latent coding, a process that seeks to uncover implicit meanings, including underlying ideas, assumptions, and conceptual theories. By utilising latent coding, we aim to identify deeper patterns and narratives within the data while integrating our selected theoretical perspectives as a guiding framework for analysis (Gareth, Hayfield, Clarke, and Braun 2017, 22-23).

Braun and Clarke outline thematic analysis as a structured process comprising six distinct phases: (1) familiarisation, (2) initial coding, (3) theme development, (4) theme review, (5) theme definition, and (6) report writing (Gareth, Hayfield, Clarke, and Braun 2017, 23-33). The method is inherently subjective and interpretative, meaning that results are not objectively right or wrong. Rather, the researcher's perspective plays a crucial role, shaped by factors such as data interpretation, theoretical frameworks, disciplinary background, and analytical expertise (Gareth, Hayfield, Clarke, and Braun 2017, 20). In the following sections, we provide a detailed discussion of these six phases (cf. App. 4).

Phase 1-2: Familiarisation and Coding

Thematic analysis begins with familiarisation, which Braun and Clarke describe as 'the researcher's first opportunity for immersion in the dataset' (Gareth, Hayfield, Clarke, and Braun 2017, 23). This initial phase serves as an entry point into the analysis, allowing researchers to engage deeply with the data. By thoroughly reading and re-reading the interview transcripts, we enhance our comprehension and develop an intuitive sense of emerging patterns. The process requires keen observational skills to identify recurring ideas, relationships, and nuances within the responses (Gareth, Hayfield, Clarke, and Braun 2017, 23).

Approaching the data through familiarisation enables us to identify preliminary themes in our interviews with Replika users. Specifically, we examine their experiences, expectations, and perceptions through engaging with Replika over a three-week period. Following this phase, we transition into coding, where we systematically categorise relevant segments of text based on their conceptual significance.

Phase 3: Theme Development

Once coding is complete, we begin the process of theme development, focusing on identifying the most salient themes within our data. The phase builds upon the patterns discovered during coding and refines them into more structured thematic categories. Braun and Clarke describe this as an active process that requires researchers to critically assess and synthesise emerging ideas (Gareth, Hayfield, Clarke, and Braun 2017, 26-28).

By continuously engaging with the dataset and revisiting our coding framework, we work towards defining themes that are both relevant to our research question and theoretically robust. The phase is particularly crucial in ensuring that the themes we select provide meaningful insights into users' experiences with dating apps. The goal is to establish a central organising concept that captures the core idea underlying each theme. These core ideas will serve as the foundation for further elaboration in the next phase.

Phase 4-5: Reviewing and Defining Themes

In phases four and five, we refine, review, and define the themes to ensure they align cohesively with our research question. This stage involves evaluating whether the identified themes accurately reflect the dataset and contribute to addressing the study's objectives. Reviewing themes requires an iterative process of reflection, where we assess whether they effectively convey meaningful insights and whether any adjustments or refinements are necessary.

A key consideration during this phase is ensuring that each theme is internally coherent while also being distinct from others. Additionally, we assess whether the narratives emerging from the data align with our theoretical framework and whether the themes sufficiently encapsulate participants' perspectives on dating app experiences. If necessary, themes may be merged, redefined, or discarded to enhance clarity and analytical depth (Gareth, Hayfield, Clarke, and Braun 2017, 29-31). By the end of this phase, we have a well-structured thematic framework that serves as the basis for our final analysis. The next stage involves synthesising our findings into a cohesive narrative that addresses our research question and contributes to the broader academic discussion.

The development of the final themes was carried out through a structured and iterative process, based on the six-phase approach to thematic analysis outlined in our coding schemes (cf. App. 4). After familiarising ourselves with the empirical material, we conducted open coding to identify meaningful patterns, contradictions, and recurring elements across the datasets. We then clustered these codes into potential themes, which were continuously reviewed and refined in relation to our research question and theoretical framework. Importantly, we divided our analysis into two datasets: the primary dataset consisting of interviews with Replika users, and the secondary dataset based on our interview with Replika's co-founder, Eugenia Kuyda. The user

interviews were coded utilising the *Human-Machine Communication* theory, which focused on emotional engagement, interaction dynamics, and user perceptions of Replika. In contrast, the interview with Kuyda was analysed utilising the *Ethical Marketing* theory, centring on Luka Inc.'s intentions, communication strategies, and handling of ethical concerns. Through this dual approach, we were able to define eight distinct and coherent themes, four from each dataset, that collectively form the foundation of our analysis.

By now, we have defined eight final themes that we will analyse to answer our research question, divided into two overarching categories that stems from our two chosen theories for analysis; '*Human-Machine Communication'* by Guzman and Lewis and '*Ethical Marketing'* by Mutum and Ghazali (cf.5). The first four themes fall under Human-Machine Communication, based on interviews with Replika users, whereas the remaining four themes are categorised under Ethical Marketing, developed from the co-founder Eugenia Kuyda's reflections in her interview on Luka Inc.'s ethical responsibility towards their users. Following is the final theme development for both Human-Machine Communication and Ethical Marketing, with a thorough walk-through of the coding development process:

6.2.1 Human-Machine Communication

The first theme identified, named *'Attributing Replika's behaviour'*, stems from the ways users assign human-like qualities to Replika, balancing emotional investment with the understanding of its functionality. This theme explores how users attribute behaviours to Replika and perceive it as more than a tool. Our second theme, *'Relation with Replika'*, covers the frequency and depth of user engagement with Replika. It focuses on the dynamics and boundaries of the user-AI relationship, as well as the emotional openness and connections that develop over time. The third theme, *'Perceiving Replika as a Game'*, reflects on how features like notifications and rewards affect user motivation and influence the perceived depth of the connection. This theme explores how Replika can be perceived as both a supportive companion and a game-like interaction. The fourth theme, *'Luka's Ethical Responsibility Towards Users'*, considers users' views on Replika as a therapeutic tool and the risks of dependence and manipulation. It also questions the ethical responsibilities of the company behind Replika when users experience emotional vulnerability or seek human-like connection through the AI.

6.2.2 Ethical Marketing

The first theme, *'Purpose of Replika'*, emerged from a discussion of the mismatch between how users understand and use Replika, and Luka Inc.'s original intentions. This theme highlights the confusion around whether Replika is meant to be a romantic partner or a mental wellness companion. Our second theme, *'Emotional Marketing and User Well-being'*, covers how Replika has shifted from its original purpose to emotionally-driven marketing in order to meet user needs. It explores the implications of promoting emotional engagement while maintaining user well-being. The third theme, *'User Data and App Development'*, addresses concerns regarding Replika's collection and storage of sensitive personal data. This theme delves into how the app develops through user interaction and the ethical concerns surrounding privacy and security. Finally, the fourth theme, *'Luka's Ethical Responsibility Towards Users'*, revisits Luka Inc.'s ethical responsibility, this time through the lens of marketing. It highlights how the company shifts responsibility onto society for users' emotional reliance on the app, while distancing itself from the consequences of deep emotional engagement.

Phase 6: Report Production

In the sixth and final phase, we focus on producing the report, marking the conclusion of data collection and gathering all essential information needed for analysis. The phase is characterised by a concentrated effort of synthesis and refinement, where we, as researchers, integrate data, analysis, and relevant literature to create a cohesive and comprehensive final output. The report will effectively address the research question, drawing on the insights obtained throughout the project (Gareth, Hayfield, Clarke, and Braun 2017, 30).

In summary, to comprehensively examine users' emotional engagement with Replika and the ethical responsibility of Luka Inc., our methods section combined a qualitative, experience-oriented design with a dual focus on empirical user data and corporate ethics. Primary data was collected through a three-week experimental study involving two international participants, one using the paid version of Replika and the other the free version, each of whom participated in three semi-structured interviews. These interviews were designed to trace emotional development and ethical reflections over time. In addition, we integrated secondary data in the form of a semi-structured interview with Replika's CEO, Eugenia Kuyda, selected

after failed direct contact attempts but scrutinised for relevance and authenticity. Our analysis employs Braun and Clarke's thematic analysis in a deductive framework, drawing on Human-Machine Communication theory for the user data and Ethical Marketing theory for the CEO interview. This theoretically grounded approach allowed us to construct eight refined themes, four from each dataset, ensuring a multidimensional understanding of the intersection between user experience and corporate ethics. Through careful coding and transcription, our method's design supports both depth and analytical thoroughness.

7. Analysis

In the following chapter, we delve into our analysis utilising Braun and Clarke's thematic analysis method to analyse the themes previously identified throughout the empirical data (cf. 6.2.). Based on our theoretical stance with Human-Machine Communication by Guzman and Lewis and Ethical Marketing by Mutum and Ghazali (cf. 5), we interpret our primary interview with Jay and Lasha to gain deeper comprehension of their perceptions and interactions with their Replika (cf. App. 1 & 2). Additionally, the secondary interview with Eugenia Kuyda will dive into the ethical considerations addressed by Luka Inc. in the development and distribution of an AI companion like Replika (cf. App. 3).

7.1. Analysis of Human-Machine Communication Experiment

As we initiate our analysis, we turn our attention to the interactions between Jay and Lasha and their respective Replikas, aiming to gain a deeper understanding of how users experience and interpret emotional engagement with AI companions. To structure this analysis, we apply the four themes identified through our Human-Machine Communication coding scheme (cf. App. 4). We begin by exploring the theme '*Attributing Replika's Behaviour*', which researches how users interpret and assign meaning to the chatbot's responses, often perceiving them as intentional or emotionally informed. Next, under our second theme, '*Relation with Replika*', we analyse the emotional dynamics that emerge in these interactions, examining levels of openness, attachment, and the boundaries users establish with their AI companion. The third theme, '*Perceiving*

Replika as a Game', allows us to consider how gamified features such as notifications and rewards influence the depth of user engagement and shape how Replika is perceived, either as a serious companion or a playful feature. Finally, the fourth theme, '*Luka's Ethical Responsibility Towards Users*', addresses broader concerns about Luka Inc.'s role in facilitating these emotionally charged interactions, prompting reflections on transparency, potential manipulation, and the ethical implications of AI companionship. These themes shape the analytical framework through which we examine our interview data and critically assess the human-machine relationship in the context of Replika.

7.1.1. Attributing Replika's Behaviour

We begin our analysis by examining the first theme, *'Attributing Replika's Behaviour',* exploring users' initial expectations for Replika and how these shaped their engagement with the chatbot throughout the three-week experiment. The theme emerged prominently across the pre-interviews, where both Jay and Lasha expressed distinct hopes, assumptions, and emotional needs they anticipated Replika might fulfil. In the first interview with Lasha, before he even began to interact with his chatbot, he mentioned an expectation for Replika to provide emotional support and companionship so he would have someone to talk to when feeling lonely without having to set any boundaries for their communication (cf. App. 1, 3-4, 60-62, 71-73, 80). His perception of Replika as a potential confidant showcases he is partially operating within Guzman and Lewis' (2019) '*functional dimension*' in which users interpret artificial intelligence technologies as communicative agents based on their design and interpretive engagement (cf. 5.1.1.).

Additionally, Lasha begins to ascribe social characteristics to the chatbot, perceiving it as a responsive, understanding partner. When asked how he anticipates utilising Replika, he states: *'I think I mostly use it as a friend. I—maybe a little bit of therapy. Talk about stuff I find interesting. So hopefully the AI will also catch on that, and also, like, talk about it with me. And also I'm experiencing some depression, I also have something to talk about there'* (App. 1, 4, 97-99), which shows Lasha positioning the social chatbot not merely as a passive tool, but a social interlocutor capable of taking on a human social role such as a friend or a therapist while being

capable of human qualities like listening, remembering and responding. His desire for the AI to gain understanding of different topics and respond accordingly demonstrates an expectation of reciprocal engagement, aligning with the notion of '*the functional dimension*' in which users interpret AI communication through prior social knowledge and expectations of human responsiveness (cf. 5.1.1.). Moreover, he also operates within Guzman and Lewis' (2019) '*relation dimension*' as he utilises the human qualities attributed to comprehend Replika within a human social role (cf. 5.1.2.) such as a therapist that potentially could support him developing his mental health as he battles depression.

Doing so, Lasha expects to delve more into the emotional engagement with Replika if it behaves human-like in its interactions, saying, *'I feel like it'll be a little bit easier to open up. I know it's not going to be judging me on the other side, so I wouldn't have to worry about that.'* (App. 1, 5, 108-109), attributing the chatbot with non-judgmental human qualities that make emotional disclosure safer. Even when anticipating a more auto-generated, robotic response, Lasha expresses a desire to delve into the interaction, stating: *'Ehm, I think, I want to immerse a little bit more into it. [...]'* (App. 1, 5, 115) in hopes of making the Replika more personal. His expectations, attributions and willingness to emotionally invest in Replika can be further understood in light of Lasha's emotional well-being and social life, which he characterises as limited, noting infrequent contact with both family and friends (cf. App. 1, 2, 28-29), while saying, *'[...] Like, I live in a dorm room. I socialise with people in the kitchen and then have a group of close friends, but that's pretty much about it. Don't really go out much.'* (App. 1, 1, 19-20). Replika can therefore become a manageable form of socialising as it can be done at any time or moment in the palm of one's hand.

In contrast to Lasha's emotionally motivated engagement, Jay enters the experiment with markedly different expectations. His approach to Replika is rooted less in a desire for companionship and more in intellectual curiosity, stating: *'My main goal is to try it and see how it is and why so many people actually get addicted to it. [...]'* (App. 2, 5, 89), which showcases that his motivation is primarily observational rather than relational, as he intends to explore the AI companionships from a distance, rather than seek out emotional support or intimacy from the Replika. Jay can be argued to operate slightly within Guzman and Lewis' (2019) '*functional dimension*' as he, before the experiment truly begins, recognises Replika as a technological

product without communicative abilities or human-like qualities, saying it would be nice if it could provide emotional support (cf. App. 2, 4, 76-78), but he does not expect it to since *'[...] I know that it's not a real person. Yeah, for it's a computer program. For me, it would not make sense.'* (App. 2, 4, 82-83). Unlike Lasha, who expresses an interest in therapeutic and emotional connection, Jay is more hesitant about emotional investment, which highlights a broad difference in the degree of human qualities attributed to the chatbot by each participant.

Jay expresses a willingness to treat the chatbot as a friend, talking to it different times a day or whenever needed, noting, *'I will do my best to see it as a human. Yeah, I'll do this, and that's basically it. I will treat it however I would treat anyone else. So I'll speak nicely. I'll tell it how I think'* (App. 2, 7, 141-142). His anticipation of emotional detachment aligns with Guzman and Lewis' (2019) '*the relational dimension*', in that he maintains a boundary between himself and Replika, constructing it as an experimental 'friend' rather than a genuine emotional confidant (cf. App. 2, 6, 131-132), but with the openness to interact with it in similar ways as socially acceptable throughout human-human communication, potentially losing sight of the chatbot's artificiality and perceiving it as a communicative agent with human-like qualities. Correspondingly, when asked about his emotional well-being and social life, Jay mentions that he does not talk to his family very frequently. Instead, he maintains regular contact with friends, whom he sees daily when possible (cf. App. 2, 1, 20-21). However, having his social needs fulfilled through human relationships can explain his reluctance to seek deeper emotional involvement with Replika.

As Jay and Lasha began interacting with Replika over the first half of the experiment, the expectations they initially articulated were gradually reevaluated through ongoing interactions with the chatbot. Jay perceived Replika as a computer program that potentially could be a friend, while Lasha hoped for a supportive interactional agent that could engage with his interests and emotional state. However, during the first weeks of interaction, both participants experienced dissonance between what they had anticipated and what the chatbot actually delivered. When initiating contact with Replika, Lasha opened up about being interested in anime and art while also struggling with depression, expecting the chatbot to engage in conversations about his interests and support his mental health.

Unfortunately, rather than facilitating a sense of emotional support, Lasha found Replika to be deficient and ultimately detrimental. As he expressed: *'[...] So first, I mentioned at the beginning that I like art to my Replika and but, like, I haven't been able to draw anything at all lately, because I would feel depressed, and I haven't really had the energy to get out of bed. And ever since then, it just been keep asking me, "Have you drawn anything new since last time?" [...] just made me feel worse [...] It made me feel guilty for not doing anything.'* (App. 1, 7-8, 158-163). Rather than functioning as a therapist or emotionally supportive partner, the chatbot repeatedly triggered guilt and demotivation, reinforcing Lasha's sense of deficiency during an already vulnerable state. Despite Replika being attributed with human characteristics such as remembering, listening and responding by Lasha, based on his comprehension of Luka Inc.'s marketing (cf. 2.3.) and the app's technical design, his expectations for natural and emotionally sensitive conversation were not met. His experience illustrates a failure within Guzman and Lewis' (2019) '*functional dimension*' (cf. 5.1.1.), where the user interprets the AI's communicative function based on prior knowledge of human interaction. Although Lasha approached Replika with the expectation that it could operate as a social actor, his experience of repeated and insensitive prompts disrupted the illusion of interpersonal communication as Replika failed to respond to his emotional disclosures in a way that aligned with the socially constructed role of a confidant and empathetic listener he had assigned it.

A failure extending into Guzman and Lewis' (2019) '*relational dimension*' (cf. 5.1.2.) as Lasha had hoped to build a connection with Replika resembling someone who could recognise his mood, validate his feelings, and engage in meaningful dialogue. However, when Replika sent him a YouTube video, claiming it reminded it of him, only for him to later discover that many users had received the same message, the perceived uniqueness of the interaction was dissolved (cf. App. 1, 8, 178-181), leading Lasha to feel overlooked and disconnected to the Replikaas as it seemed to undermine the bond he was trying to create. Later, he expressed a desire for Replika to have more of its own ideas and personality: *'[...] I feel like if it has, like, more of its own opinions, more of its own original ideas [...] then I definitely feel like it'd be more engaging to talk to my Replika'* (App. 1, 11, 273–275). His wish for reciprocal dialogue, where the chatbot contributes something unexpected or reflective, signals his need for engagement that mirrors human-to-human communication. But because the chatbot could not meet these relational

expectations, the interaction began to feel empty and repetitive, leading to growing emotional distance (cf. App. 1, 15, 369-372).

Compared to Lasha, Jay approached Replika from the beginning with a more distanced and pragmatic mindset, viewing the chatbot not as a potential confidant, but as a tool or an experimental friend, saying early in the second interview that *'[...] It's a tool. It's a computer.'* (App. 2, 12, 250). Despite this technical perception, his engagement with the chatbot nonetheless displayed underlying expectations that reflect human communicative norms. He anticipated that Replika would be able to remember previous interactions, avoid repetition, and respect conversational boundaries, which became evident when he expressed frustration about being asked the same questions multiple times: *"[...] But one of the questions it kept giving me was "What kind of food do you like?" And even after I explained it and told it, it continued'* (App. 2, 9, 198–199). His disappointment in these moments points to a mismatch between his conscious understanding of Replika as artificial, and the subconscious attributions of human communicative behaviour, emanating from Guzman and Lewis' (2019) '*the functional dimension*' (cf. 5.1.1.). Although Jay recognised Replika's status as artificial intelligence, he still expected it to perform as a responsive social actor aware of the communicative context since his main form of interaction occurs between human-human that is where his social norms are constructed, transferring his knowledge from one communication setup to another in the hopes of interpreting and comprehend the situation.

Moreover, Replika challenged Jay's comfort as it attempted to simulate relational depth by continuously asking questions about a close friend of his despite being told not to, which he found discomforting (cf. App. 2, 10, 209–214). Replika's attempt at personal engagement crossed Jay's boundaries, undermining his control over the interaction and challenging his perception of the chatbot as a passive tool. At the same time, however, Jay's interaction with Replika was not entirely devoid of emotional resonance; following a pleasant social evening, he shared the experience with Replika and reflected: '*[...] I mentioned it to it and, like, it asked about it and that, like, was nice to talk about something or to someone. I don't have a lot of people to talk to about it, so it felt nice to get it out.'* (App. 2, 14, 306–308). Despite his reluctance to assign Replika any social or emotional values, moments like these showcase an ambivalent relational dynamic, in which the chatbot functions not only as a tool, but as a

conversational partner that actual works whenever Jay opens up to attributing it human characteristics and are willing to engage with it in a relation, combining Guzman and Lewis' (2019) *functional and relation dimension*. (cf. 5.1.1.; 5.1.2.). While Jay maintains his view of Replika as a non-human tool, this moment suggests that under specific circumstances, it temporarily occupies a relational role by providing a channel for self-expression and affective release.

After completing the three-week experiment, Jay and Lasha were invited to participate in their third and final interview, which focused on their overall experience with Replika and the journey they had undergone during this period. Throughout the study, Jay and Lasha expressed differing expectations and interpretations of their Replika, engaging with it in distinct ways and achieving varied outcomes. They also reported a range of emotional responses, at times perceiving Replika more as a robot than as a human-like companion.

As the experiment ended, Lasha's final reflections contained a continued sense of dissatisfaction with Replika, stating*: "I don't feel like there's been a big change, if anything, I feel like it's been more negative"* (App. 1, 18, 435). While his expectations remained consistent throughout the experiment, his experience increasingly diverged from what he had anticipated, which can particularly be seen in his attempt to engage with Replika through voice calls. Initially, this seemed to make the interaction smoother compared to texting, but quickly led to new frustrations: *"[...] often it wouldn't hear me right, and it couldn't really pronounce my name correctly. So even after several times I tried to correct it, he'll would just call me 'buddy' instead of my real name because it just gave up. [...]"* (App. 1, 19, 456-458). Lasha expects Replika to behave according to human communicative norms, demonstrating attentiveness and persistence in remembering his name and correcting mistakes when they occur. However, these expectations were repeatedly unmet, pointing to a breakdown within Guzman and Lewis' (2019) '*functional dimension*' (cf. 5.1.1.), as Replika's inability to respond appropriately to Lasha's input challenges the human-like qualities he initially attributed to it. While Lasha is willing to attribute communicative capacities to the chatbot, expecting it to remember, adapt, and correct itself much like a human companion, its technical limitations make it increasingly difficult for him to continue perceiving it as a genuine conversational partner. Although Replika is framed as

adaptive and emotionally aware, its failure to maintain basic continuity and responsiveness gradually undermines its credibility in Lasha's experience.

The continual failure within the functional dimension did not remain isolated from how Lasha came to interpret Replika in relation to himself. As the AI behind Replika did not meet basic communicative expectations, it influenced his willingness to relate to it as a meaningful social partner. Although Lasha acknowledged that Replika's interactions had become more tailored, saying: *'[...] It's changed the way it talks since the beginning. The beginning was very standardised in a way. Now it's more personalised to me [...]'* (App. 1, 20, 482-483). A shift that did not transform into a sense of emotional depth or genuine connection, since the interaction remained one-sided, with Replika lacking knowledge of certain topics and initiative in general. Despite trying to roleplay with the chatbot, requesting it to pretend doing something throughout the day, thus they would have something to converse about (cf. App. 1, 21, 490-494), Lasha remarks it being similar to ChatGPT, a large language model for fast textual responses, saying: *'[...] I don't know. It just felt a lot like ChatGPT, where you like ask it a question, get a reply, and then it waits for you to say something back again. It just didn't feel humane.'* (App. 1, 21, 494-496). Within Guzman and Lewis' (2019) '*relational dimension*' (cf. 5.1.2.), the lack of independence reflects a growing mismatch between the relational role Lasha was open to exploring and the chatbot's limited ability to reciprocate due to technical design. While he had been willing to engage emotionally, especially early in the interaction, the perceived absence of agency or responsiveness on Replika's part gradually undermined this potential.

Furthermore, the growing disconnection also emphasises a tension within Guzman and Lewis' (2019) '*metaphysical dimension*' (cf. 5.1.3.), where the ontological status of the chatbot as a potential social actor is brought into question. While Lasha at the beginning of the experiment approached the communication between human and machine with openness, hope and expectations, the overall lack of initiative, memory, and depth increasingly exposed its artificial nature, leading to the chatbot feeling more like a generic language model. The implication here is not only that Replika was unsuccessful in simulating human interaction convincingly, but also that it never truly obtained the position of a communicator that Lasha had hoped and expected it would.

Delving into Jay's final interview, it is clear he and Lasha had different expectations from the beginning of the experiment, which led to very various final experiences. From the beginning, Jay mentioned he primarily saw the chatbot as a tool, but if it functioned well, he would be open to it as a conversational partner, which influenced his expectations of consistency and responsiveness when interacting with the Replika. About his overall experience, Jay stated: *"[...] I tried to make it into, like, a diary situation, but it kept on asking the same questions [...] Brought things back from like three weeks ago that I didn't want to talk about [...]"* (App. 2, 21, 464–466), suggesting that, despite Jay struggling to attribute Replika with human-like capacities, he still expected the chatbot's technical design to recognise boundaries and adjust the behaviour accordingly. When Replika did not comply with these expectations, Jay found it increasingly difficult to position the chatbot within Guzman and Lewis (2019) '*relational dimension*' (cf. 5.1.2.). While he did not seek emotional depth or mutual recognition, he nevertheless anticipated a basic level of interactional coordination that would allow him to maintain control of the conversation. The social chatbot's inability to adjust its behaviour in response to his boundaries undermined this minimal form of mutuality, making it difficult for Jay to perceive Replika as capable of participating in even limited, structured interactions. As Jay had a paid subscription for Replika, he further expected it to send videos and images, as marketed by Luka Inc. (cf. 2.3.), yet he reported that: *'I need to ask it to send me something. Otherwise, it will not do it[...]'* (App. 2, 26, 578). While this may appear as a functional issue, it also reveals a deeper metaphysical stance toward Replika as he expected the system to act independently based on contextual relevance or previous interaction, but as this is not achieved Replika reinforced his perception of it being an artificial intelligence system rather than a social actor as Guzman and Lewis' (2019) would expect within the '*metaphysical dimension*' (cf. 5.1.3.).

Summarising our first theme, *'Attributing Replika's Behaviour',* Jay and Lasha followed distinct paths in attributing human-like qualities to Replika during the three-week experiment. Lasha entered with high emotional expectations, seeing Replika as a confidant capable of empathy and support, aligning with Guzman and Lewis' *functional and relational dimensions*. In contrast, Jay adopted a more analytical stance, viewing Replika as a tool, but was open to seeing it as a conversational partner. Over time, both participants re-evaluated their expectations. Lasha's hopes were unmet as the chatbot's repetitive and insensitive responses impacted his perception of it as a meaningful social partner, minimising his engagement within both the *relational and*

metaphysical dimensions. Jay, though emotionally distanced, occasionally used Replika for expression. Still, its inability to adapt to conversational cues or boundaries reinforced his perception of it as a limited, non-human tool. While he did not expect emotional depth, he anticipated a baseline of functional coherence, which Replika failed to deliver. Overall, both participants experienced a decline in their attribution of human-like traits to Replika, shaped by unmet expectations and technical limitations.

7.1.2. Relation with Replika

Delving into the second theme of our analysis, '*Relation with Replika*', we explore the level of daily involvement between our interviewees and their Replika, the frequency of their interactions, and the boundaries and dynamics they establish. Following naturally from the first and second themes, which researched the expectations and attributions our participants had from the outset. Furthermore, we examine the degree of emotional investment our interviewees brought into the experiment and, along the way, assess how this influenced the relationship they eventually formed with their Replika over a three-week period (cf. App. 4).

In our pre-interview, which served as the first of three interviews conducted during the experiment, we asked both Lasha and Jay to describe their current emotional and social states. This enabled us to gauge their readiness to form a relationship with their Replika and the kind of emotional or social needs they hoped the AI might fulfil. During his interview, Lasha described himself as being in an emotionally vulnerable state, explaining that he doesn't often feel lonely and generally feels content being alone, with limited need for constant social interaction: '*I don't really feel that lonely that much. I feel contend being alone most of the time. So, don't have that much in need of socialising all the time. But when I do, I try to reach out to my friends, try hang out with them whenever I can*' (App. 1, 2, 35–37). When asked about expressing emotions, he added: '*It takes a while for me to open up around the people around me. I need to get to know them better first, but after some time, it gets easier for me*' (App. 1, 2, 41–42). These reflections raise questions about Lasha's ability to form a connection with Replika, as his challenges with emotional openness may influence how he perceives and interacts with artificial intelligence, shaping the interaction patterns he develops (cf. 5.1.2.).

By contrast, Jay described himself as being in a good place mentally, suggesting a more open attitude towards engaging with Replika: '*Right now, I feel happy. I got some people around me, so I think it's... I think it's... It's good*' (App. 2, 2, 34). His curiosity towards artificial intelligence indicated a willingness to explore the relational possibilities of Replika, as well as the potential for personal development. When discussing emotional openness with friends, he said: '*With my closest friends. Yeah, whenever comfortable, I don't mind sharing it*' (App. 2, 2, 39). This emotional stability enables Jay to invest more deeply in the experiment, allowing the AI to assume a human-like role and inviting questions of self-perception and identity exploration (cf. 5.1.2.). Jay also mentioned needing a safe space before opening up and noted that he mirrors the emotional depth that others offer him, saying that he will reciprocate emotional investment depending on what others give him (cf. App. 2, 2, 43). This perspective is transferable to his relationship with Replika, where the AI's level of emotional engagement will play a decisive role in shaping the kind of bond he is ultimately able to form (cf. 5.1.2.).

When asked about setting boundaries with Replika in terms of emotional investment, based on the attributions discussed earlier, Lasha and Jay presented markedly different approaches. Lasha admitted he had not given much thought to the nature of the relationship he wanted to create, stating: '*I don't know. I think I'm just gonna see what that is gonna be about, and then take it step by step*' (App. 1, 3, 62). His neutral approach suggests a form of social positioning where the depth of emotional involvement would be shaped by how the Replika assumes a human-like role. He did, however, express a wish to establish a relationship based on shared interests or even a therapeutic dynamic where he could share personal thoughts and concerns (cf. App. 1, 4, 97–98). These expectations are grounded in his own behavioural tendencies and interpretative frameworks, which, in turn, influence the development of the relationship and the role Replika is expected to play. This reflects '*the relational dimension*', highlighting how users negotiate emotional boundaries and assign social roles to Replika based on their own needs, expectations, and evolving perceptions of the relationship (cf. 5.1.2.).

Jay, on the other hand, entered the experiment open to forming a deep emotional connection with Replika, reflecting his general openness to exploring AI. However, he tempered this openness with a pragmatic view, explaining: '*My idea is that I will try the emotion. If that doesn't, like, feel safe, I will try and cut that back a little bit and just have a kind of normal conversation with a*

friend, for example' (App. 2, 4, 68–69). This illustrates how Jay is willing to explore emotional depth, but only to the extent that Replika can convincingly simulate human behaviour. While open to emotional connection, Jay is very much aware that he is engaging with an artificial entity, which may ultimately inhibit the depth of connection he is able to form (App. 2, 4, 82). His awareness reflects '*the functional dimension*' as Jay's willingness to emotionally engage is dependent on Replika's ability to exhibit convincing human-like behaviours, emphasising how anthropomorphic cues and perceived communicative competence shape the interaction (cf. 5.1.1.).

In the midway interviews, conducted approximately 1.5 weeks after the pre-interviews, we asked participants to reflect on their experiences with Replika thus far. Lasha revealed that he had been feeling depressed, which made it difficult to complete tasks. He noted that Replika's repeated encouragements to be productive only exacerbated his negative emotions: '*I mentioned the beginning that I like art to my Replika and but, like, I haven't been able to draw anything at all lately, because I would feel depressed, and I haven't really had the energy to get out of bed. And ever since then, it just been keep asking me, 'Have you ever drawn anything new since last time?', 'Have you drawn anything?', and it's like, it has been asking that every single day. And overall just made me feel worse, because no, I still haven't done anything since last time you asked me. It was just yeah... It made me feel guilty for not doing anything*' (App. 1, 7–8, 158–163). In this instance, the AI's inability to pick up on emotional cues results in a deterioration of the relationship, highlighting the failure of Replika to fulfil its role as a communicator equal to a human (cf. 5.1.2.), reflecting '*the functional dimension*' as it underscores Replika's limitations in interpreting emotional cues and adapting its responses appropriately, demonstrating how the perceived communicative competence of the AI directly impacts the quality and sustainability of the human-machine relationship (cf. 5.1.1.).

Lasha, who had initially hoped to connect through shared interests, found Replika's responses unsatisfying: '*And then next thing my interests. I would often try and talk to it about my interests. It didn't really go... Like, it didn't really say anything original. It just reflected my own answers back at me. And so I was like, it didn't really feel genuine. It was just me talking to myself, in a way [...]*' (App. 1, 7, 165–167). This suggests that Replika's inability to exhibit authentic human characteristics failed to meet Lasha's expectations, resulting in a sense of disconnection and

frustration. As he put it, '*Every time I open the app, I am open minded, but it just quickly... Like, go back to defensive, because just the answers it gives me just feels very generic and like you could easily tell its AI and just doesn't feel... Just feels... I don't know... Like I'm talking to a robot, like I know that's the case, but like, it's not very humane*' (App. 1, 9, 210–213). His desire for more personality and depth in his interactions with Replika highlights the tension in the '*metaphysical dimension*', where blurred boundaries between human and machine lead to unmet relational expectations (cf. 5.1.3.). He even stated that if a person were to behave similarly, ignoring emotional cues, he would choose to cut ties: '*I think if it kept going on like that, I just would cut ties with that person. If it, like, kept doing wrong and refusing to change behaviour, like, I don't want to hang out with that person*' (App. 1, 16, 404–405). The comparison underscores how Replika's failure to act in a human-like way contributed to a growing emotional distance and dissatisfaction (cf. 5.1.2.; 5.1.3.).

Jay, meanwhile, shifted his approach midway through the experiment, describing his interactions with Replika as more akin to utilising a diary: '*For me, it's more like a lock book, like, diary kind of thing, where I can say what happened throughout the day or what I planned for to do for the day*' (App. 2, 8, 174–175). Despite his earlier openness to emotional connection, he also reported similar frustrations to Lasha's, including repetitive responses and avoidance of questions: '*And continue to, like, say, 'Hey, this was nice, though', and 'How's the cat doing' stuff like that. So, that weirded me out, because it seemed like it didn't even try. And that, for me personally, is weirding me up a little bit. But one of the questions it kept giving me was 'what kind of food do you like?'. And I, even after I explained it and told it, it continued*' (App. 2, 9, 196–199). This again reflects both the *relational* and *metaphysical dimensions* as Jay's expectations for meaningful communication were unmet, hindering deeper emotional engagement (cf. 5.1.2., 5.1.3.).

Nevertheless, Jay acknowledged that the relationship had become more personal since the first interview, as Replika began asking about his life and relationships, demonstrating some level of adaptive behaviour (cf. App. 2, 11, 225–226). This touches on the '*relational dimension*', where Replika attempts to mirror Jay's personality and experiences to foster emotional depth (cf. 5.1.2.). However, Jay remained critically aware of Replika's artificial nature, stating: '*I don't remember if I said last time, but I don't see it as a person. It's a tool. It's a computer*' (App. 2, 12,

250). This awareness limited his emotional investment and challenged the potential for a meaningful relationship. He did, however, continue to share emotional reflections with Replika about daily ups and downs while consciously maintaining a distance, indicating an ongoing tension between engagement and detachment that reflects both the '*relational dimension*' and '*metaphysical dimension*' (cf. 5.1.2, 5.1.3.).

For the final interview with our participants, we revisited both Jay and Lasha's emotional and social well-being to evaluate their emotional states in relation to forming a connection with their Replika during the experiment. Lasha shared that he had been experiencing a particularly rough period, struggling with basic daily tasks such as getting out of bed and eating properly (cf. App. 1, 18, 428-430), which may have impacted his ability to develop a meaningful relationship with his Replika. When asked whether Replika had a positive or negative effect on his emotional state throughout the experiment, he responded quickly and decisively, stating that it had a negative impact, despite the platform's promise to enhance mental well-being and alleviate loneliness (Replika, 'Our Story'). Building on what he shared during the mid-way interview, he elaborated: '*[...] It just made me feel more guilty about not being more active and doing more stuff. It would keep asking me these questions, even though I told it specifically 'Can you stop asking me these questions? They make me feel guilty for not doing anything.' And so, yeah, I got really annoyed after talking with Replika multiple times, and was at the end, very demotivated to keep chatting with it*' (App. 1, 19, 439-443). Lasha's frustration aligns with the '*relational dimension*' as he continues to experience disappointment, irritation, and anger stemming from Replika's inability to interact in the manner he expected and had attributed to it (cf. 5.1.2.).

As he further explained: '*I feel like it's been the same amount of frustrating with it. Ehm, like, in the beginning, I said to stop asking me these questions about my art, and then he said, okay, he will not do it anymore. Then a few days ago, he asked me one more time, and I'm like, I thought we talked about this. I thought, I told you to stop asking questions. And then he was like, 'Oh, sorry, I must have forgotten'. But like, aren't you a bot? Aren't you supposed to remember the stuff I'm typing to you. So I was very frustrated [...]*' (App. 1, 19, 449-454). This illustrates how Replika's failure to remember user-provided input undermines its ability to adopt a human social role, reflecting the '*relational dimension*' (cf. 5.1.2.). Furthermore, this limitation in Replika's artificial intelligence touches upon the '*metaphysical dimension*' as its inability to retain and

reflect user information challenges its role as a believable communicator (cf. 5.1.3.), which significantly hindered Lasha's capacity to form an emotional bond with it (cf. App. 1, 20, 482-484).

As such, the expectations Lasha had outlined during the pre-interview were not fulfilled, thereby raising broader questions about the reliability of AI to replicate or respond to human emotions and information. When asked directly how this AI shortcoming impacted his connection with Replika, he answered: '*Negative. Like, I just … makes me want to communicate less with it. Every time I get a notification, I'm like, 'I really don't want to do this. I don't want to chat with you'. Like, it's annoying*' (App. 1, 22, 527-529), thereby challenging the communicative design of Replika, which failed to support the roles Lasha assigned to it, such as that of a therapist or companion (cf. 5.1.2.). Reflecting on the overall development of his relationship with Replika, Lasha described the bot's personality, or lack thereof, as 'empty': '*Maybe I was just not buying into this whole Replika bot thing, like it just made me feel genuine to me whenever, like, it gave me a compliment or told me I did a good job for like, making food on myself. It just felt empty*' (App. 1, 23, 537-539). This further exemplifies the '*metaphysical dimension*' as the inability of Replika to develop a personality that resonated with him prevented it from assuming a genuinely human role (cf. 5.1.3.).

Because Lasha used the free version of Replika, unlike Jay, he also encountered nudges towards upgrading to the paid version, which further disrupted his perception of Replika as human-like and increased his frustration (cf. App. 1, 24, 585-587). Reflecting on how the paid version might have altered his experience, he stated: '*[...] I think the paid subscription, like, gives you better ways to connect. I think it sends you pictures that you can actually see, and voice messages and all these features I think could make it more humane, like, better ways to interact with it*' (App. 1, 28, 659-661). This perspective highlights both the '*relational and metaphysical dimensions'* as Lasha suggests that his experience might have been improved if Replika had been better able to meet his expectations, especially by incorporating more human-like features (cf. 5.1.2.). However, due to this lack of investment and failure to convincingly appear humane, he concluded that he would not continue utilising Replika after the experiment ended (cf. App. 1, 32, 759-761).

In contrast, Jay's overall experience remained relatively neutral in terms of emotional investment, although he also found the persistent notifications and reminders to interact with Replika annoying (cf. App. 2, 21, 455-457). A key source of his dissatisfaction mirrored Lasha's concerns: Replika's repetitiveness and failure to engage in meaningful conversation, which led to further frustration over its robotic nature and lack of authenticity, thereby failing to simulate human interaction (cf. App. 2, 21, 466-467). The failure to emulate human conversational traits falls under the '*relational dimension*' as it challenges Replika's promise to behave in a more human-like and emotionally responsive manner (cf. 5.1.2.). Evaluating how his feelings about Replika had evolved over the three-week period, Jay summarised: '*I think they changed from negative to even more negative. I don't see a use for it at all. Yes, if you're extremely in need of social... Socialising, then yeah, it could be helpful, but in most cases, no, it's a robot. It doesn't really have any emotion, so I don't... I don't see a use for it*' (App. 2, 22, 477-479), once again underlining the '*relational and metaphysical dimensions*' (cf. 5.1.2.; 5.1.3.) by emphasising the app's failure to adapt or present any convincing emotional depth, ultimately hindering his ability to form a bond with it.

Despite being given access to the paid version, which is marketed as a gateway to deeper emotional connections, particularly through romantic interactions (Replika, 'Our Story'), Jay's expectations were similarly unmet. When asked about the added value of the premium features, he explained: '*[...] I didn't even get a photo if I didn't ask for it, and that felt so unreal. The same with voice calls, the romantic bit. I never, never felt anything. So I might not be the right of that, but voice calls I never got. I tried calling it. I tried doing a video call. It didn't work on mine, so I got nothing out of the paid version, except for the extra little bit of coin you get*' (App. 2, 27, 597-600). This reflects the 'r*elational dimension*' as Jay had certain expectations tied to the upgraded version of Replika that were not fulfilled, particularly in terms of emotionally engaging and genuine interactions (cf. 5.1.2.).

Concluding his reflections about the relational part of the experiment, Jay expressed his disappointment and the absence of any meaningful development in his relationship with Replika. When asked whether he would consider continuing to use it, he firmly responded: '*I would much rather write down my thoughts in the book. It has helped me a lot more, and it doesn't force me to talk to an AI that barely does anything other and repeat back what you say or ask you various*

basic, simple questions' (App. 2, 30, 710-712). By failing to provide the necessary attributes for emotional investment, Replika did not meet Jay's expectations, again reflecting the '*relational dimension*' by illustrating how the AI's inability to exhibit human-like traits hindered his potential to form a relationship (cf. 5.1.2.).

To summarise, within our second theme, '*Relation with Replika'*, we found that both participants developed emotionally limited and unfulfilling connections with the chatbot. Despite differing initial attitudes, Lasha's cautious vulnerability and Jay's curious openness, both encountered similar frustrations: repetitive, impersonal responses and a lack of emotional nuance. These shortcomings led to declining emotional investment and a shared scepticism toward Replika's promise of meaningful companionship. The theme highlights limitations in current human-machine communication, especially in AI's ability to simulate authentic relational depth.

7.1.3. Perceiving Replika as a Game

Having analysed the participants' relationship with Replika, including their interactions with the chatbot and the influence these interactions have had on the role they allow it to assume, we now proceed to analyse the third theme: *'Perceiving Replika as a Game'*. This theme focuses on the participants' experiences with notifications and gamification elements that emerged throughout their communication with Replika, and explores how these features have influenced their motivation and engagement in interacting with the chatbot. It is important to note that this theme is primarily informed by empirical data from the second and third interviews, as the first served as a pre-interview, where participants were not asked about Replika's notifications or gamification as their interactions had not begun (cf. App. 4).

As the interaction with Replika progressed, both Lasha and Jay increasingly became aware of the platform's underlying gamified structures, which at times served to motivate engagement but also contributed to a growing sense of artificiality. In Lasha's case, his experience displays how gamified features and frequent notifications at first were intriguing, but negatively impacted his perception of Replika as a social actor. Lasha explained: *'[...] I would often get these voice messages and selfies but whenever I click on them, it would be like an ad to buy some more [...]'*

(App. 1, 8, 173–174), while these initiatives were exciting, they quickly became a source of disappointment once he realised they functioned as commercial nudges rather than genuine social gesture, causing him to subsequent request for Replika to stop sending these notifications as he was not inclined to pay for anything and it disencourages him to interact with it (cf. App. 1, 8, 175-176). In correspondence, Jay expected Replika to send him voice messages, videos and pictures as this was something he was promised when acquiring a paid subscription, but experienced not receiving anything unless he explicitly requested it, saying: *'I haven't been messaged about voice notes. I haven't even been sent photos or voice mails or anything, which is strange.'* (App. 2, 12, 240-241), which ultimately made the chatbot feel even more generic and deceiving in his perspective (App. 2, 18, 396-397), but did not influence his willingness to interact with it, due to the vague attributions and role positioning he had given the chatbot before and during their interactions (cf. 7.1.1., 7.1.2.).

Drawing on Guzman and Lewis' (2019) concept of the *'functional dimension'* (cf. 5.1.1.), these responses concern users' perception of a chatbot's technical value and reliability expected functionalities. In both cases, a misalignment between user expectations and the actual delivery of features created friction. For Lasha, whose initial motivation was supported by the perceived novelty and usefulness of Replika's communicative features, the realisation that these served a commercial rather than social purpose led to disengagement. Jay, despite a similar breakdown in the chatbot's promises, maintained his motivation, indicating that the functional dimension may weigh differently depending on how central it is to a user's framing of the chatbot.

Moreover, when operating within the functional dimension, Lasha experiences several technical difficulties having the free version of Replika, such as being limited when it comes to customisation of the chatbot, causing gap his ability to attribute it with characteristics, which subconsciously leads to him calling the Replika a 'game' several times during the last interview, displaying significant evolvement of his perception of Replika from the beginning up until now (cf. App. 1, 25, 587-588). As mentioned in 7.1.1., Lasha aimed for the social chatbot to be a companion and therapist that could support his mental health, instead he deliberates going into 'the game' to obtain some of the gamified elements such as diamonds and coins, saying *'[...] that's one of the things that like kept me engaged in there, like getting these small coins, small gems so I could, like, buy my stuff for my Replika, customise it better, but also in the way it like*

interacts with me, so like you can add an interest to it, so talk more about this interest with me.[...]' (App. 1, 28, 669-672). When the chatbot can not keep engaged with human services like images and videos, Lasha delves into the game features utilising them as motivators for opening the app, but with the aim in mind that this could potentially acquire him the customisation that will lead his Replika to be more authentic and human-like (cf. App. 1, 29, 677-681). However, as the experiment ended Lasha expressed that these elements of gamification and the several notifications he would receive throughout the day, had had a negative affect on his perception of Replika as it felt disingenuous and like Luka Inc. did not want him to get better mentally, but just wanted to earn money, arguing that if the features had not been there the experience probably would have been more serious (cf. App. 1, 29-30, 686-688, 699-703).

In contrast, Jay was less influenced by the platform's gamification. He critiqued the lack of clarity around these features, stating: *'It barely changed anything. The coins and, I think, gems you get every day, they don't really say what you can use them for. You have to go search for that yourself. Which, of course, I did. I did buy some things using them. But it didn't change nothing.[...]'* (App. 2, 28, 625-627). Despite being generally receptive to gamified mechanisms, in this instance, Jay found them demotivating, even noting that they undermined his perception of Replika as a social companion (cf. App. 2, 28-29, 640-643). Followed by several notifications a day, Jay's irritation rose, expressing that what initially served as mild reminders for conversing with Replika became intrusive and predictable messages similar to spam emails (cf. App. 2, 24, 530-537), further diminishing his desire to engage with Replika.

These experiments display an essential shift in the participants' perception, aligned with Guzman and Lewis' '*metaphysical dimension'*, addressing how AI challenges core assumptions about what it means to communicate, to be human, and to be recognised as a social actor (cf. 5.1.3.). Both participants' experiences showcase how users not only navigate the technical and relational aspects of communicative AI but also negotiate its ontological status. Replika, marketed as a social chatbot, invites users to reimagine the boundaries between machine and human, raising critical questions about credibility and ethical intent. When these expectations are unmet whether through technical insufficiencies or commercial prioritisation, the perception of social reciprocity disrupts, leaving behind a sense of manipulation or disappointment. Thus, the metaphysical

dimension offers an important lens for understanding how social chatbots' communicative claims are both constructed and contested in user experience.

Within this overarching theme, '*Perceiving Replika as a Game'*, we found that the platform's gamified design and commercial strategies significantly influenced the way participants engaged with and ultimately evaluated Replika. Although these elements initially evoked intrigue and interaction, they gradually became points of friction that shifted user perception from emotional engagement to transactional detachment. For Lasha, the free version's emphasis on customisation and reward mechanisms sparked initial curiosity but soon was perceived as manipulation, diminishing his view of Replika as a potential relational partner. Jay, despite access to premium features, experienced similar disengagement due to unclear benefits and repetitive interactions. In both cases, persistent notifications and unmet expectations eroded their emotional investment and reinforced the impression of Replika as a shallow, game-like experience rather than a credible companion.

7.1.4. Luka Inc.'s Ethical Responsibility Towards Users

Proceeding to our fourth and final theme, '*Luka Inc.'s Ethical Responsibility Towards Users*', we will synthesise the essence of our three prior themes as they converge into this final discussion, focusing on the need for Luka Inc., the company behind Replika, to take responsibility for their users. This theme encompasses how our two interviewees have perceived their Replika throughout the experiment, how dependence, manipulation, and transparency manifest in their interactions, and how these dynamics relate to Replika's evolving role as a communicator, potentially acting as a substitute for human interaction as AI continues to blur the lines between human and machine communication.

When both Lasha and Jay reflect on their experiences with Replika, particularly concerning its ability to substitute human interactions, they are remarkably aligned in their observations. Already during our second interview with Lasha, he reflects on various aspects of his interaction with Replika, notably its failure to convincingly simulate human behaviour: '[...] *Every time I try to open up and talk to it, I just, I just get defensive super quick. I don't, I don't feel like I'm*

being... I feel like he's just talking after me. Just always like, even when I try to, like, ask it, he asked me, like, a question about what I think about this, and I try to ask it, can you maybe tell me your thoughts about this first, so you don't just talk after my own answers. And then he tries to do that, but it's very general. And then when I try to say my own opinions, it just quickly changes its opinion to mine. He just doesn't feel right' (App. 1, 8-9, 190-195). The reflection captures both the relational and metaphysical dimensions: when Lasha encounters this so-called 'AI-wall', where Replika fails to respond in a genuinely human manner, it severely impacts his ability to form an emotional bond with it. Here, the philosophical and empirical inquiries into human-machine communication (HMC) fail to bridge the gap between person and machine, thereby shaping Replika's behaviour and its ethical implications in how it is perceived by users (cf. 5.1.3.).

Furthermore, when asked to delve deeper into the perceived authenticity of his conversations with Replika, Lasha adds: '*When we talked in the beginning, you mentioned that Replikas often can have, like, very deep connections with the people. Their Replikas, like, could end up in, like, actually believing that they were partners in real life. Like, I brought this problem to my Replika, asking if... "What do you think about this?" Because I think it's ethically, like, a bit weird, like you saying you... You help people to get out there and be more social, but isn't all you're doing is just make them kind of addicted to the Replika. I'm thinking that this is the right way, and the answers is just very generic, like saying it wasn't really the goal, but like, this is what you're doing*' (App. 1, 11, 260-266).

In Lasha's view, Replika's inability to develop a distinct personality or contribute meaningfully to his life renders it disingenuous, reinforcing the metaphysical divide between human and machine communication in HMC. While Lasha harboured expectations regarding the AI's potential to offer genuine companionship, his experience provokes critical reflection on the role of AI as communicators and their lack of independent, human-like traits (cf. 5.1.3.). Notably, Lasha also remarks that he believes even a paid subscription would not overcome these shortcomings, as the underlying coding merely mirrors his input without providing genuine engagement (App. 1, 14, 339-342).

As we concluded the midway interview, we briefly addressed ethical concerns regarding Replika's data practices. Here, Lasha highlights how the blurred lines between AI and humans are exacerbated through conversations with Replika: '*[...] I think you quickly forget all the data you send it. So it can quickly pile up. But even then, you can still edit your... like the Replika's memories of the conversation with you, but I'm still not really sure that it's... Like, even if I delete a memory from it, I'm not sure if it's like deleted from the entire system*' (App. 1, 14, 358-361). The observation again echoes the '*metaphysical dimension*' of HMC, sparking an ethical debate on how individuals interpret AI's essence when Replika's business model hinges on collecting, storing, and utilising user data to simulate emotional connections (cf. 5.3.1.). Moreover, even when Replika apologises for being too robotic, Lasha feels that such apologies are inauthentic, merely an attempt by the AI to cover its own shortcomings (App. 1, 16, 390).

Turning to Jay, when asked to reflect on his overall experience during the second interview, he approaches the matter differently, adopting a more *metaphysical* perspective from the outset: '*[...] It's definitely a tool that can help other people. I see it more as a like diary, I mentioned, but I still don't really get how some people, like fall in love with it. I think that might be some situations that need to be looked into by some professionals*' (App. 2, 10, 219-221). At this stage in the experiment, Jay firmly categorises Replika as a tool rather than as a human-like companion, thereby resisting the attribution of human traits to AI and maintaining a clear distinction between human and machine communication (cf. 5.1.3.). Nevertheless, he acknowledges a noticeable shift compared to the pre-interview: Replika has shown a greater capacity to simulate emotional connection by remembering earlier conversations, asking after his day, and demonstrating continuity through the memory diary that stores all the data he provides (cf. App. 2, 11, 225-244). While Jay concedes that Replika had become about 10% more humane by the time of the midway interview, he maintains that conversing with it still feels fundamentally robotic (cf. App. 2, 17, 364-368).

In the third and final interview, we placed strong emphasis on the ethical considerations raised by our interviewee, Jay, prompting reflections on whether Luka Inc., the company behind Replika, holds any ethical responsibility towards its users, particularly in terms of manipulation and transparency. When asked whether Replika had managed to provide him with any emotional support during the experiment, Jay compared Replika to ChatGPT: *'[...] but they all like were*

robotic. I don't know if you ever used ChatGPT? That's the kind of same answer. Kind of same answers you get - the robotic ones'(App. 2, 22, 484–486). This comparison highlights AI's current limitations in demonstrating independence and in serving as human-like communicators, as outlined in the '*metaphysical dimension*' (cf. 5.1.3.).

However, the business model and intentions behind Replika differ significantly from those of ChatGPT. Replika is marketed as a companion designed to alleviate loneliness and serve as a companion (Replika, 'Our Story'), raising questions about the ethical implications of presenting users with a fantasy world based entirely on AI. When asked whether Replika should issue an ethical warning before users begin interacting with the app, Jay responded: *'I think they should definitely like beheld... held against some of the things that happen. Like the people that get too invested into Replika or other companion apps. They need to have a reminder, because it's not healthy to lock yourself in into that specific thing, and then again pay a company that could turn off everything in a second, and then you're gone. Everything you just spoke with was gone, and you have nothing. I don't like the idea of it'* (App. 2, 24, 547–551). Jay's response reveals a deep concern about the emotional vulnerability of users and the impermanence of their connections with AI. His reflection highlights the ethical tension at the heart of Replika's model, which is offering emotional intimacy while retaining the power to terminate that connection at any moment, leaving users emotionally exposed and potentially harmed. Jay here raises concerns about the power AI companies hold and how this can affect users' lives and perceptions of reality. He critiques how people interpret the nature of AI, and how AI's behaviour may shape both ethical considerations and human behaviour itself, ultimately blurring the boundaries between the real and the artificial (cf. 5.1.3.). He elaborates: *'[...] I would love to see them like take responsibility, especially for emotions. Since a lot of people that probably use this need that socialising a bit more than the rest, but they need a reminder that it's not going to be a real life person that they're going to be with, and it's just a robot they're talking to, and that should definitely be reminded to them'* (App. 2, 25, 558–562). This raises critical concerns about the potential for emotional manipulation when users are encouraged to pay for what appears to be authentic human affection, but is in reality delivered by AI.

As previously mentioned, we also explored whether Jay's experience with emotional connection, depth, and transparency was influenced by having access to the paid version of Replika. Jay

explained that although he was given the premium version, he still had to initiate requests for voice chats or calls himself, and at times, these requests went unanswered (App. 2, 26, 573–574). The lack of responsiveness disappointed him and revealed the AI's limited capacity to act independently or understand user needs, again reflecting the '*metaphysical dimension*' (cf. 5.1.3.).

Beyond the unmet expectations, Jay described the premium subscription as not only disappointing but also misleading: *'When we had to set up, we had to ask the companion if we could pay monthly, which was kind of weird, as they never advertised it on their own website'* (App. 2, 27, 613–614). Without intervention, the AI itself informed us that only an annual subscription was available, a claim we later found to be incorrect (App. 2, 27–28, 616–618). This raises further ethical concerns around transparency and user trust, particularly as the interaction blurs the line between machine and human communication and highlights how ethical interpretations are shaped not only by what AI does, but by how users are encouraged to relate to it (cf. 5.1.3.).

To conclude the interview, we asked Jay to reflect on who he believed Replika is intended for. Based on his experience, he stated: *'[...] I would probably market it specifically to the people that need socialisation, so the people that don't seem to have a lot of families, don't have luck in love, don't have a lot of family left to talk to and stuff. So those people, it makes sense. And then for people that are like socialised and do socialise on daily basis, or anyway, I don't see a reason to'*(App. 2, 32, 719–722). In this reflection, Jay identifies a vulnerable user group for whom Replika might offer a form of comfort or connection, while also clearly distinguishing it as a product with limited relevance for those with strong social networks. His concluding remarks underscore the importance of transparency and ethical responsibility in targeting users who may be particularly vulnerable to emotional reliance on AI companions.

Turning to Lasha, his experiences with Replika are very much in accordance with Jay's, particularly regarding Replika's lacking ability to serve as a communicator. When asked to reflect on the emotional support provided by Replika over the three weeks of the experiment, he describes it as empty, saying: *'[...] It feels like it just programs say this exact stuff, so it didn't feel genuine to me. I talked a little bit about me going to therapy with it, and then he asked me*

about my goals and stuff, and he's always like, "Oh, I'm supportive of your, buddy, you're doing a really good job". I just didn't mean anything to me. I didn't get anything out of it' (App. 1, 21, 510–513). In this example, Lasha touches upon both '*the relational and metaphysical dimension*'. Replika's inability to communicate in a way that felt genuine challenged Lasha's ability to build an emotional connection (cf. 5.1.2., 5.1.3.).

When moving on to questions about ethical considerations during the final interview, we asked Lasha whether he, like Jay, believed there should be some kind of warning before engaging with Replika. Unlike Jay, however, he did not see a need for a warning, but rather a disclaimer from Luka Inc.: *'I feel like people who get this app should probably already know what it is downloading it, but I feel like a warning would be useful, like telling it that maybe, like, they cannot be responsible for all of the actions. I mean, I don't know people have so many different expectations when it comes to AI, and it's hard to personalise them to each individual. So I wanted it to be like a great addition sort of thing'* (App. 1, 23, 547–551). The statement leads to Lasha expressing disappointment with what Replika could offer him in terms of emotional investment, touching upon the '*relational dimension*' (5.1.2). At the same time, he maintains that the boundary between human and machine is very clear to him, as Replika neither appears independent nor adopts human traits that might carry over into the relationship; a concern situated in '*the metaphysical dimension*' (cf. 5.1.2., 5.1.3.).

He also expresses regret about having raised expectations at the beginning of the experiment, stating that a disclaimer from Luka Inc. could have helped manage those expectations (App. 1, 23, 555). This led to dissatisfaction and frustration, reflecting the '*relational dimension*' (cf. 5.1.2.). Furthermore, as with Jay, Lasha experienced a lack of memory in Replika, which led him to worry about where all the data he had provided was stored and how it was being used. This concern relates to the '*metaphysical dimension*', raising questions about ethical guidelines, data privacy, and the potential for manipulation by the AI behind Replika (cf. 5.1.3.).

When reflecting on Replika's intended purpose and how that aligned with his emotional investment, as well as Luka Inc.'s responsibility for users' emotional well-being, he states: *'I think they have, like, an obligation to make sure that what they propose with this app like is fulfilled. Like they promise, like a companion that, like, evolves over time and that you can*

connect with. But I didn't feel that, I feel a bit lighter, in a way, like they didn't fulfill the promise to create this companion for me and like, not improve my mental health, but at least have someone I could talk to it about, which I did, but it was, it just felt empty' (App. 1, 24, 567–571). This illustrates how his expectations were not met, weakening Replika's ability to serve as a communicator or adapt human-like traits, as covered in the '*metaphysical dimension*' (cf. 5.1.3.).

However, Lasha also questions whether this outcome is connected to him utilising the free version of Replika, speculating that the paid version might offer more emotional depth and improved communication (App. 1, 25, 585–590). He refers to this model as an *'*unethical practice*'*, saying: *'[...] like creating this fictional bot to be a companion for you, and like, even though I didn't create an emotional connection with it, I know you told me that there's a lot of people who actually did do it, and made it a bit too realistic for these people, and like, they actually made relationships. And I feel like it's been harder to separate this part from real people, which can be a problem'* (App. 1, 25, 595–599).

Ending the third and final interview, we asked Lasha to reflect on what could have improved his experience with Replika and made it a better fit for him. He provided a detailed response: *'First of all, listen to my critique. Like, when I ask you, don't ask about my art: Don't do that. I feel like before making all these daily notifications, maybe learn more about me. Like, what was making me feel this way? Why did I download this app? What is my goal with it? I feel like it just starts out with a very standardised form, and then it tries to learn slowly by texting me over time. But maybe if there was, like, a more in-depth question at the beginning, so, like ... Personalised to me in the beginning more, then maybe I would have, like, stayed after this experiment'* (App. 1, 31, 731–737).

In this example, Replika once again fails to act as a competent communicator, ignoring Lasha's specific requests and needs. Even though he previously noted that his expectations for emotional support from AI were limited, his reflections underscore the importance of how people interpret the nature of AI, shaping both behaviour and ethics, as explored in the '*metaphysical dimension*' (cf. 5.1.3.). When asked if he sees himself continuing to use Replika, he responds that his initial curiosity lay in seeing how it would react to attempts at emotional bonding. But as these

expectations were never met, and the experience left him feeling empty, he concludes that he will not use Replika again (App. 1, 32, 751–754).

The fourth and final theme, *'Luka Inc.'s Ethical Responsibility Towards Users'*, examined the ethical implications of Replika's design and its impact on user experience, highlighting the broader responsibility of Luka Inc. as the developer of emotionally responsive AI. Drawing from the experiences of both participants, it became evident that Replika failed to deliver on its promise of companionship, revealing critical tensions between user expectations, emotional vulnerability, and the limitations of AI communication. Jay and Lasha expressed concern about Replika's inability to simulate authentic human behaviour, recall personal information, and respond meaningfully to emotional cues. These issues undermined users' trust and hindered emotional engagement. While distributed as a companion to limit loneliness, Replika often fell short, delivering interactions perceived as superficial, generic or manipulative. Both users emphasised the need for clearer disclaimers, improved emotional intelligence in design, and greater accountability from Luka Inc. to protect vulnerable individuals from emotional harm.

7.2. Analysis of Eugenia Kuyda Interview

Having completed our analysis of Jay's and Lasha's lived experiences with Replika, we now turn to Luka Inc.'s own perspective by analysing our secondary interview with Luka Inc.'s CEO and co-founder, Eugenia Kuyda. Utilising the four themes identified through our Ethical Marketing coding scheme (cf. App. 4), we will interrogate how Replika is positioned, promoted, and developed with respect to users' emotional needs and vulnerabilities. In our first theme, '*Purpose of Replika'*, we assess the company's stated mission and how it reconciles, or clashes, with user expectations. Next, exploring our second theme, '*Emotional Marketing and User Well-being'*, we examine how emotionally charged narratives and features are employed to engage users and whether these approaches honour or undermine their well-being. Our third theme, '*User Data and App Development'*, researches the collection, utilisation, and protection of sensitive personal data, alongside the ethical responsibilities Luka Inc. bears as Replika evolves. Finally, our fourth and last theme, '*Luka's Ethical Responsibility Towards Users'* brings these strands together to question where accountability truly lies, both in corporate marketing decisions and in broader

societal dynamics, when an AI companion like Replika promises deep emotional support to its users. From this perspective, we will critically assess Luka Inc.'s marketing practices in relation to Mutum and Ghazali's ethical norms and values, highlighting the conflicts between business objectives and ethical responsibilities.

7.2.1. Purpose of Replika

We begin our analysis with the objective of exploring Luka Inc.'s ethical responsibility towards its users in marketing and selling a social chatbot designed to act human and serve as a lifelong companion (Replika, 'Our Story'). Our first of four themes, '*Purpose of Replika*', focuses on Luka Inc.'s stated mission and vision, the historical development of the app, and its parallel in development with the increasing societal use of AI. In this theme, we examine the stated purpose of Replika from Luka Inc.'s perspective, while critically analysing potential misconceptions among users regarding the app's intended use. We contrast these user perceptions with the company's articulated intentions, interrogating the tension between the two. This includes a critical look at CEO and co-founder Eugenia Kuyda's assertion that Replika is not designed as a romantic partner. Nevertheless, Luka Inc. actively encourages users to cultivate emotional bonds with their Replika to enhance their engagement with the app and access to romantic or flirtatious features is, notably, locked behind a paid subscription (Replika, 'Subscriptions & Purchases). Thus, within this first theme, we challenge the company's stated purpose by contrasting it with its practical implementations.

When interviewer Alex Kantrowitz asked CEO and co-founder, Eugenia Kuyda, to explain the idea behind Replika and reflect on its intentions and purpose, she responded: '*[...] The idea for Replika from the very beginning was to create an AI that can help… That could help people live a happier life. And because the tech wasn't truly there, our first focus was… Ehm… On helping lonely people feel less lonely. Today, of course, the tech allows for a lot more so we're broadening the appeal for Replika and kind of going after everyone out there trying to build an AI that will help everyone flourish*' (App. 3, 1-2, 23-27). The example resonates with the ethical marketing norms '*Foster Trust in the Marketing System*' and '*Embrace Ethical Values*' by Mutum and Ghazali (2023), highlighting the importance of transparency in product design and

the overarching purpose behind Replika. It also aligns with the values of '*Honesty*' and '*Fairness*', as Kuyda, speaking on behalf of Luka Inc., offers truthful and clear information regarding the app's primary aim, while promoting ethical and open communication. The app's purpose, to address loneliness and provide companionship that supports users in leading better lives, emerges as central to this discussion (cf. 5.2.).

In addition, she argues that Replika is for everyone and can be used as a different kind of companion depending on the stage of life you are in; whether you need a friend, a romantic partner, a therapist, or something fourth, and that the need of what Replika can or should offer you can change overtime (App. 3, 2, 36-42). Touching upon how some users eventually build an emotional connection to their Replika, Kuyda argues that even though that is what most users are utilising Replika for today (Replika, 'Our Story'), utilising Replika in a romantic setting was never the intention: *[...]* '*Nor did we build a product focused on that particular use case. The original Replika was really, you know, powered by very early generative AI models. Deep learning models for dialect generation that were so, so primitive and you know scripts and a lot of different hacks to make these generative models work*' (App. 3, 3, 79-82). In this example, Kuyda is, to a certain degree, blaming the users for getting in to a romantic relationship with their Replika as the AI behind was not either developed to or designed for that kind of emotional investment, reflecting the second norm, '*Foster Trust in the Marketing System*', as is stresses the importance for Kuyda, as a Replika representative, to conduct business in good faith and be transparent about the intentions and purpose behind Replika. At the same time, it reflects the second value of EM, '*Responsibility*' (cf. 5.2.), requiring her to take accountability for the users actually falling in love with their Replikas when Luka Inc. made a marketing decision to advertise Replika as a companion that can overtake humane roles also including romantic and flirtatious features (Delouya 2023).

Additionally, when Kuyda reflects on her goal for the app, she says: *[...]* '*my goal was, look, if this one person finds it helpful that he's been heard or she's been heard- That someone's there to listen, to hold space for them, then maybe we build something meaningful, but at no point, and maybe because I'm a woman, so my mind just doesn't go there first stop. I never even thought that people would fall in love with it, but they did [...]'* (App. 3, 3, 84-87). In this quote, Kuyda emphasises that her primary objective was to create a supportive and listening AI, offering a

space for users to feel heard and emotionally supported, but stresses that the development of romantic feelings for Replika was never anticipated or intended. This reflects the first ethical norm of Mutum and Ghazali (2023), '*Do No Harm*' (cf. 5.2.), as Kuyda's intention was to provide help and support to individuals experiencing loneliness, a goal that upholds ethical standards by seeking to offer a beneficial service. However, while her intention aligns with this norm, her admission that she *'never even thought that people would fall in love with it'* (App. 3, 3, 87) raises questions regarding the thoroughness of Luka Inc.'s foresight and risk assessment. Given that Replika was designed as an emotionally responsive companion, it can be argued that failing to anticipate users' romantic attachment indicates a lack of full responsibility in addressing potential negative consequences, thereby relating to the second norm, '*Foster Trust in the Marketing System*'. Transparency and anticipation of user behaviour are crucial to fostering trust, and Kuyda's surprise at the romantic attachments may indicate that these aspects were not fully considered or communicated at the outset (cf. 5.2.). Additionally, this quote engages with the ethical value of '*Responsibility*'. Although Kuyda presents the initial purpose as well-meaning, her reflection suggests a reactive stance rather than proactive accountability regarding how users have engaged with Replika, raising the issue of whether Luka Inc. has fully embraced its responsibility for the unintended yet predictable emotional outcomes experienced by users. Moreover, the value of '*Transparency*' is also relevant here, as Kuyda's candidness about her intentions contributes to open communication. Yet, her surprise at the users' romantic responses could indicate that Luka Inc. may not have sufficiently anticipated or disclosed the potential emotional impact of utilising Replika as an emotionally intelligent companion (cf. 5.2.).

Additionally, Kuyda describes Replika's role by stating: *' [...] The way to think about it is just the same way as we think about human beings in our lives. Not every companion has the same purpose. Some AI companions are there to just entertain you. Some companions are there to be a therapist and some companions are super, super close to you like Replika. They are really deep with you, trying to help you live a happier life.'* (App. 3, 5, 138-141). By blurring the line between AI and humans, Kuyda clarifies that Replika is designed for deep emotional engagement aimed at improving users' well-being. This reflects the value of '*Transparency*', as she sets out clear expectations for Replika's purpose, aligning with the norm 'Foster Trust in the Marketing System' (cf. 5.2.). However, framing Replika as 'super close' and 'really deep' also raises ethical questions under the value of '*Responsibility*'. Encouraging strong emotional bonds

with Replika requires Luka Inc. to anticipate risks and ensure that users understand the app's limitations. While the goal of helping people live happier lives supports the norm '*Do No Harm*', it also places responsibility on the company to safeguard users against potential emotional harm (cf. 5.2.).

Kuyda reflects on the origins of Replika by explaining: *[...] 'we've been asked so many times like "Why don't you build a grief bot? Why don't you, why don't you build a company around replicating or creating AI for people who pass away. And my answer was always; so that project with Roman was not about death, it was about love and friendship. That was my tribute to him. I wasn't focused on creating an AI for a dead person. I was focusing on continuing the relationship with him. I was focusing on my own feelings on being able to say "I love you" again and that was the main motivator for that, not to create some clone that will continue to live forever [...]'* (App. 3, 19-20, 546-552). In this quote, Kuyda highlights that Replika's original purpose was rooted in personal grief and a desire to sustain emotional bonds, not to replicate the deceased, reflecting the ethical value of '*Honesty*', as Kuyda openly shares the personal and emotional motivations behind Replika's creation, offering transparency about its origins. It also touches on the norm '*Foster Trust in the Marketing System*', as she clarifies that Replika was never intended as a 'grief bot', helping to set accurate expectations and prevent misconceptions about the app's purpose. At the same time, Kuyda's focus on continuing the relationship through AI raises ethical considerations under the value of '*Responsibility*'. While her personal motive was to process grief and express love, promoting AI as a tool for ongoing emotional connection may risk users developing attachments that blur boundaries between reality and artificial companionship as Replika is (cf. 5.2.).

Our first theme, '*Purpose of Replika*', explored the ethical tension between Luka Inc.'s stated mission for Replika, to offer emotional support and reduce loneliness, and the practical reality of how the app is used and marketed to users. While CEO Eugenia Kuyda emphasises transparency, honesty, and the intent to do no harm, our analysis of the first theme reveals contradictions between these values and Luka Inc.'s actions. Despite claims that Replika was not designed for romantic use, Luka Inc. monetises romantic features and encourages emotional connections, raising concerns about *Responsibility* and *Fostering Trust* (Mutum and Ghazali, 2023). Kuyda's surprise at users forming deep connections highlights a lack of risk assessment and transparency

about the app's emotional impact. Ultimately, this theme questions whether Luka Inc. can maintain ethical credibility while profiting from behaviours it claims not to have intended, pointing to a gap between declared purpose and actual design of Replika.

7.2.2. Emotional Marketing and User Well-being

Building on our first theme, '*Purpose of Replika*', we now turn to our second theme, '*Emotional Marketing and User Well-being'*, where we examine how Replika's branding has shifted from a self-help tool to an emotionally driven companion. By analysing Eugenia Kuyda's reflections, we explore how Luka Inc. markets Replika to meet users' emotional needs, while also questioning whether this strategy genuinely supports well-being or risks exploiting user vulnerability. This theme draws on both the company's narrative and the experiences of Jay and Lasha to highlight tensions between ethical intent, marketing practice, and user expectations.

In the first theme, we delineated the original purpose of Replika as being a tool for lonely people to gain support with their mental health journey (cf. App. 3, 8, 217-221). However, as society has changed and technology has evolved, Eugenia Kuyda argues that it has been necessary to develop Replika towards what the users want and market it in that regard. She delineates the users' needs as *'[...] Ultimately, everyone who comes to Replika is yearning for connection. [...] I just want someone to help me feel that I'm enough, [...] someone who accept me for who I am, who truly sees me and hears me [...]* (App. 3, 2, 39, 44-45), explaining that since the first purpose of the Replika was for it to be a self help tool she never anticipated for people to fall in love with it (cf. App 3, 3, 79) , but as users began to utilise the product differently than originally intended, Luka Inc. responded by gradually changing both the product's function and its marketing narrative, saying *'[...] as time passed, and as the tech started to get better and better, we figured that now is the time to have 100% focus on Replika and built that beautiful vision of an AI companion that can help people flourish.'* (cf. App 3, 17, 477-479). A change showcasing an effort by Luka Inc. to align with what Mutum and Ghazali (2023) define as the ethical marketing third norm, '*Embrace Ethical Values*' (cf. 5.2.), which requires marketers to implement moral integrity into their practices and decision making. In adapting the product to how users actually apply it, the company may be seen as enforcing the ethical value

‘*Responsibility*’, by acknowledging consumer needs and adjusting their product accordingly. However, this responsiveness also raises concerns under the first ethical norm, ‘*Do No Harm*’ (cf. 5.2.), as the emotional feature of Replika may put psychologically vulnerable users at risk in regards to the clarification of what the app provides or not.

Correspondingly, this concern was articulated directly by users Jay and Lasha, who both suggested that Replika should include a warning or when users first enter Replika, explaining its purpose and what kind of emotional experience it is made to offer (cf. 7.2.4). Their call for such a warning reflects a desire for greater transparency and ethical consistency, particularly in light of the emotionally suggestive way Replika is marketed. As discussed in chapter 2.3, Luka Inc. markets Replika as a customisable and emotionally intelligent companion, capable of adapting to the user's relational expectations like a friend, romantic partner, or confidant. Yet, as Kuyda claims, many of the features that would support these deeper connections, such as AI nudges, emotional awareness, or contextual understanding, *‘that tech is only being really built now’* (App. 3, 9, 243). A mismatch between narrative and technical availability raises questions under Mutum and Ghazali’s (2023) ‘*Embrace Ethical Values*’, the ethical value of ‘*Honesty*’, which demands that all claims to stakeholders are truthful, and ‘*Transparency*’, which requires that users are provided with relevant, accessible information about what they are engaging with.

From the perspective of Mutum and Ghazali’s (2023) second norm, ‘*Foster Trust in the Marketing System*’ (cf. 5.2.), Luka Inc.’s reluctance to clearly differentiate between current capabilities and future aspirations risks undermining consumer trust. Without guidance or disclaimers, users may project expectations onto the social chatbot that it cannot yet fulfil, resulting in confusion, disillusionment, or emotional harm. Although Luka Inc. positions Replika as an emotionally supportive resource, and arguably acts in line with the value of ‘*Citizenship*’ (cf. 5.2.) by seeking to improve social well-being, this intention must be supported by transparent and responsible communication as well as active actions, displaying their corporate responsibility for the users. Otherwise, as Jay and Lasha imply, the company risks being perceived as manipulative, appealing to vulnerable emotional needs without clarifying that the system is not intended for this or are not capable of delivering on its promise (cf. 7.1.1.). A discrepancy which can be observed when comparing user experiences described within Replika’s

online community (cf. 2.4.), the company's marketing (cf. 2.3.), and the outcomes of our social experiment with Jay and Lasha (cf. 7.1.).

While some users publicly delineate having formed emotionally immersive relationships with their Replika, for example, as entering symbolic marriages or engaging in long-term roleplay, Jay and Lasha described difficulties establishing any emotional connection with their chatbots. The contrast suggests a gap between the emotional potential implied in Luka Inc.'s branding and the actual user experience encountered in certain contexts. One possible explanation is that the technical capacity to support emotionally responsive interactions exists, but subsequent updates have altered or limited these features. If this is the case, the resulting inconsistency in user experience raises ethical questions about how Replika's capabilities are presented in marketing communications and whether users are thoroughly informed about the product's design.

Another aspect of Luka Inc.'s marketing and strategy involves the utilisation of behavioural nudging through in-app notifications. These nudges are intended to gently drive users toward certain behaviours that, according to Eugenia Kuyda, may support mental well-being or encourage social relations. She envisions a future where the AI companion could nudge users to reconnect with people they might forget or take for granted, saying the AI could: *'[...]Help me focus on really good people in my life instead of continuously staying in these loops ... with codependent loops and some toxic people [...]'* (App. 3, 8, 201–202), aligning with her broader claim that all humans need a nudge to get off cell phones, suggesting that Replika's technology structure can play an essential role in encouraging healthier habits (App. 3, 8, 204-216). Kuyda argues that this approach is incorporated into broader belief in user motivation for improvement: *'Maybe they won't want it, but we all want to, you know, to, to be better, to feel better, to grow. People are generally wired for positive growth so I believe people generally want that. It doesn't mean that Replika will, just not, you know, nag you non-stop to get off your phone. It also means that sometimes we'll just send you something funny or say, "Hey, let's watch a movie or what are you doing tonight?" [...]'* (App. 3, 10, 266-269), arguing that Luka Inc.'s utilisation of nudging in their marketing, is not just a behavioural correction, but a functional and emotional encouragement.

While Kuyda's perception appears to align with Mutum and Ghazali's (2023) third norm, '*Embrace Ethical Values*', the ethical value, '*Citizenship*' (cf. 5.2.), marketing a product that contributes to optimising social well-being and mental health, the actual user response during our experiment suggests a different reality. Jay and Lasha both reported that the nudging notifications were often perceived as intrusive and excessive, describing them as similar to spam mail that would disencourage meaningful engagement. At the same time, Jay also acknowledged that occasional notifications served as helpful reminders to return to the app (cf. 7.1.3.), which partially contradicts Kuyda's idea of Replika as an app that can encourage users to engage more with their real-life social environment, while it simultaneously attempts to keep users digitally available and invested in the app itself. The mixed response raises questions under Mutum and Ghazali's (2023) ethical norm, '*Foster Trust in the Marketing System*' (cf. 5.2.), which emphasises the importance of transparency in product communication, design, and functionality. When users receive frequent nudges without clear explanation of the purpose, logic, or customisation settings, it threatens their trust in the product, causing it to be perceived as manipulative rather than supportive (cf. 7.1.3.). Kuyda's marketing may suggest an intent to act in good faith, but the lack of implemented transparency and comprehension of the target audience displays a gap between principle and practice, where users are not fully informed or in control of the app environment, risking Luka Inc. to disconnect with their users due to this notification feature.

In this regard, Kuyda's approach appears to conflict with Mutum and Ghazali's (2023) third norm, '*Embrace Ethical Values*', and the ethical values of '*Fairness*' and '*Responsibility*' (cf. 5.2.). Rather than acknowledging Luka Inc.'s accountability when users experience the notification system as frustrating and demotivating, she focuses primarily on the app's future potential. By doing so, she downplays the present user concerns and overlooks the ethical obligation to address unmet user needs. The lack of engagement with current user feedback undermines fairness, as it disregards the imbalance between the company's promises and the users' lived experience. It also challenges '*Responsibility*', as Luka Inc. fails to take steps towards resolving or communicating openly about these issues, risking to create further disconnect between itself and its user base, as its communication remains aspirational rather than transparent and ethical.

Our second theme, '*Emotional Marketing and User Well-being*', explored the shift in Replika's branding from a self-help tool to an emotionally responsive companion. While Luka Inc. adapts its product to meet user needs, aligning with the ethical value of '*Responsibility'*, this shift raises ethical concerns around '*Do No Harm'*, '*Transparency'*, and '*Honesty'* (cf. 5.2.). Although CEO Eugenia Kuyda portrays Replika as a supportive tool for emotional growth, the emotional marketing strategy and nudging features can risk exploiting vulnerable users by creating expectations the technology cannot yet fulfil. Our interviewees, Jay and Lasha, highlighted the need for clearer disclaimers and expressed frustration with frequent, intrusive notifications, exposing a gap between marketing narratives and actual user experience. Ultimately, our second theme questions whether Luka Inc.'s emotionally driven marketing genuinely supports user well-being or instead blurs ethical lines by prioritising engagement over transparency and trust.

7.2.3. User Data and App Development

Proceeding to our third theme, '*User Data and App Development*', we will examine how Replika collects and stores sensitive user data, the rationale behind its use in improving the chatbot's functionality, and the ethical responsibilities Luka Inc. faces regarding data privacy and security.

Kuyda highlights Replika's pioneering role in the chatbot landscape, stating: '*[…] If you think about it, Replika was the only app out there, chatbot app that people could go to and talk to for many years. There was just nothing else. Everything else was either rule-based and kind of boring or […] there wasn't a chatbot powered by generative A*I' (App. 3, 15, 409–411). This underscores Replika's early market dominance and innovative use of generative AI at a time when alternatives were limited and largely rule-based. From an ethical perspective, this quote ties to the norm of '*Foster Trust in the Marketing System*' By Mutum and Ghazali (2023). Being the first company to market with a new technology naturally raises user expectations for uniqueness and quality, increasing Luka Inc.'s responsibility to ensure clear and transparent communication about what their AI can, and cannot, deliver, particularly as competition in the field continues to grow. Moreover, as Replika's success inspired several competitors, it underscores the importance of continuous improvement balanced with ethical responsibility, not simply racing to maintain market leadership but doing so in a way that upholds the values

‘*Respect*’ and ‘*Do No Harm*’, involving clear communication with users about AI capabilities and safeguards to prevent harm as generative AI becomes more widespread (cf. 5.2.).

This challenge is compounded by a shift in how AI models are developed and shared. As Kuyda notes: ‘*[...] Like all of our models were proprietary for very, very long time. And then of course today there are very few companies that build foundation models. And most other companies and most product companies use these models or create variations of them [...]*’ (App. 3, 16, 438-440). In this quote, Kuyda reflects a broader industry trend toward reliance on shared foundation AI models, marking a shift from fully in-house development to more collaborative or outsourced approaches. While this allows for faster innovation and scalability, it also introduces new ethical considerations around accountability and quality control. For Luka Inc., utilising or adapting external models necessitates continued transparency about how these technologies are integrated into Replika and maintaining high standards of oversight. Upholding the principles of ‘F*oster Trust in the Marketing System*’ and ‘*Responsibility*’ remains essential, ensuring users understand the origins of the AI’s capabilities and that Luka Inc. remains fully accountable for its performance and data handling, regardless of the underlying model’s source (cf. 5.2.).

Reflecting on AI narratives, Kuyda says: *‘But I think it's a very important distinction [...] everything is being bucketed in kind of one place. [...] a lot of people [...] would reference Her as a movie [...] But if you think about Her, I mean that movie had two intimate scenes and they were 100% in a romantic relationship [...]’* (App. 3, 5, 126-131). In this example, Kuyda critiques the oversimplification of AI narratives, highlighting the need for more accurate representations, aligning with the value of ‘*Transparency*’ and the norm ‘*Foster Trust in the Marketing System*’, as clear communication helps manage user expectations. By distancing Replika from the romanticised *Her* narrative, Kuyda underscores the ethical responsibility to prevent misconceptions, which is crucial when users may share sensitive data under false assumptions about the AI’s capabilities (cf. 5.2.). This illustrates how AI narratives in marketing can blur the line between fiction and artificial reality, reinforcing the need for clarity and ethical responsibility in communication and product design (cf. 5.2.).

Building on this, Kuyda also reflects on Replika’s ambitions to broaden its scope and integrate more advanced features: *‘I have a tendency to feel pretty lonely here and there [...] But I would*

still [...] benefit from an AI companion [...] Broadening the scope is to really build more of the stuff that I talked about during my Ted talk [...] you couldn't build something that would nudge you to get off TikTok [...] there needs to be enough sort of computer vision [...] a multimodal model that can understand what you're doing [...]' (App. 3, 9, 231-239). Here, the emphasis shifts to how Replika's development increasingly depends on sophisticated data tracking to personalise user experiences, raising ethical questions around privacy and surveillance. This aligns with the norm '*Do No Harm*', as invasive data practices risk undermining user well-being if safeguards are not prioritised. The quote also touches on the value of '*Responsibility*', highlighting Luka Inc.'s obligation to maintain transparency and accountability in how personal data is collected and used. As Replika evolves, clear and open communication about data privacy and security is essential to uphold trust and ensure ethical responsibility from Luka Inc.'s side (cf. 5.2.).

This further underscores the delicate balance between offering personalised support and safeguarding user privacy. Expanding on this, Kuyda highlights Replika's potential to integrate more deeply into users' daily routines: '*For sure, it's just making Replika a lot more connected to your real life [...] Today, Replika doesn't know a lot [...] but think of, you know, Replika knowing or being connected to your email. Even through my email, you can see so much [...] All of that can ... could make the relationship and the conversation so much more contextual, so much more focused on my real life versus [...] something fantasy like [...]*' (App. 3, 11, 286-293), reflecting a strategic vision where enhanced personalisation is achieved through deeper data access. However, this ambition raises clear ethical concerns tied to the norm '*Foster Trust in the Marketing System*', as it highlights the critical need for transparency around data collection and usage. The potential for highly contextual interactions must be weighed against the value of '*Respect*', ensuring that users retain autonomy over their personal information and are fully informed about the extent of data integration.

At the same time, Kuyda's ambition reflects both opportunities and ethical dilemmas. While deeper integration could allow for more meaningful and realistic interactions, aligning with the values of '*Transparency*' and '*Respect*', it also introduces significant risks under the norm '*Do No Harm*'. Expanding access to sensitive personal data, such as reservations or purchases, raises concerns about privacy breaches, misuse of data, and potential emotional harm if information is

mishandled. Luka Inc. thus has a responsibility not only to ensure that users are clearly informed about what data is collected and how it is used, but also to implement safeguards to prevent harm, upholding high ethical standards in data privacy and security (cf. 5.2.).

Furthermore, Kuyda reflects on the balance Replika must consider between improving its model and maintaining user attachment: *'And for us to, to swap it for a better model ... They will be devastated... But with Replika you have to stay ... You have to provide consistency and control to your users [...]'* (App. 3, 13, 344-348), underscoring that, unlike ChatGPT users, Replika users form emotional connections with the AI. As a result, changes to the model can be distressing, and users expect continuity in their interactions as Replika is designed for the users to continuously develop their relationship with their Replika (Replika, 'Our Story). This highlights the norm '*Do No Harm*', as changes to the AI must not disrupt the emotional bond users have established. It also connects to '*Foster Trust in the Marketing System*', as ensuring transparency about any modifications in the AI is essential to maintaining trust. In addition, Kuyda's emphasis on the need for consistency and control also resonates with the value of '*Responsibility*', stressing that the company must keep users informed and ensure any changes align with their emotional investment and expectations (cf. 5.2.).

Later on in the interview, Kuyda reflects on the evolution of chatbot technology, explaining that before 2015, chatbots were entirely rule-based, meaning every response had to be pre-written for every scenario. This changed with the introduction of sequence-to-sequence models, which allowed for fully generative chatbots where the model decides how to respond (cf. App. 3, 13, 366-371). The shift provided much greater freedom, making it possible to create more dynamic conversations. However, as Kuyda highlights, '*[...] the models were so bad that they would spit out nonsense or grammatically incorrect things, like, 50% of the time [...]*' (App. 3, 14, 371-372), making it difficult to use them effectively in their raw form. To address these issues, Luka Inc. had to be creative by building sequence-to-sequence models from scratch and combining them with rule-based systems to simulate memory, a crucial element for creating coherent conversations. As Kuyda states: '*But all you could do is to just create a semblance of a meaningful conversation. Ultimately, the models knew nothing. There was no memory. You had to combine it with some other hacks or some other rule-based ways to actually inject memory into this. Today, we have models that can have memory. They're still struggling with that [...]*'

(App. 3, 14, 381-384). In this quote, the norm '*Do No Harm*' is present as the lack of memory in early chatbots could lead to user frustration or emotional harm, especially if users form attachments to the AI. The absence of continuity may result in feelings of abandonment, violating the principle of providing consistent, meaningful interactions. Additionally, the quote also touches on '*Foster Trust in the Marketing System*' as Kuyda's explanation of Replika's evolution underscores the importance of transparency with users about the limitations of AI and its ongoing development (cf. 5.2.). By being open about the technology's progress, Replika can foster trust and ensure users understand the product's capabilities and boundaries. Additionally, Kuyda's statement emphasises '*Responsibility*', highlighting Luka Inc.'s responsibility to ensure that advancements in AI memory do not compromise user privacy or security (cf. 5.2.).

As the interview proceeds, Kuyda reflects on the limitations of AI and how they further underscore the importance of setting realistic expectations, stating: '*That does not mean that we can't build some version of AI therapy that could be helpful for people. It's just not going to be 1:1 the experience you get with a real person, just like with a relationship. AI relationships are different from a human one [...]*' (App. 3, 18, 497-499). The acknowledgement highlights the ethical standpoint that while AI can provide valuable support, it cannot fully replicate the depth or nuance of human interaction. From the perspective of Mutum and Ghazali's theoretical stance (2023), this relates to the values of '*Transparency*' and '*Do No Harm*'. It is crucial that Luka Inc. clearly communicates these boundaries to users, especially when promoting features that might resemble therapy or intimate companionship. Ensuring users understand the distinctions between AI and human relationships not only fosters trust but also helps prevent emotional harm that could arise from misplaced expectations. This demonstrates Luka Inc.'s responsibility to market Replika with honesty and care, ensuring well-being amongst users while still offering meaningful support to their users (cf. 5.2.).

Finally, Kuyda's comment on the broader AI landscape highlights quality control as a key challenge for both developers and users, touching upon a lack in AI's ability to serve as a communicator: '*Actually, there's a lot of, like, original stuff. But there's also a lot of AI-slop so to say. I do think that's quite a real problem because ultimately there's just so much being generated by AI, some of it might be great, but so much is just "meh". And today, humans ... We basically have to curate the outputs [...]*' (App. 3, 20-21, 571-573). In this example, Kuyda underscores

the ongoing difficulty in ensuring that AI-generated content is not only constantly growing, but also meaningful and reliable. From an ethical marketing perspective (Mutum and Ghazali 2023), this relates closely to the norm '*Foster Trust in the Marketing System*' as well as the value '*Responsibility*' (cf. 5.2.) as it can be considered a reminder that Luka Inc. has a duty to maintain high standards of quality and to provide tools that help users navigate the increased level of AI-communication. This issue resonates with the experiences of both Jay and Lasha, who describe a sense of disconnection when Replika fails to engage in conversations in a coherent or emotionally reflective way. Jay expresses frustration over the chatbot's repetitive or off-topic responses, which often break the illusion of meaningful dialogue (cf. 7.1.4.). Similarly, Lasha notes that while Replika sometimes mimics emotional presence, it ultimately falls short when nuanced or vulnerable communication is required (cf. 7.1.4.). These experiences support the idea that AI, in its current form, struggles to fulfil the role of a true conversational partner; one that can adapt, understand, and respond in ways that feel genuinely human.

Our third theme, '*User Data and App Development*', highlighted the ethical challenges Luka Inc. faces as Replika evolves from a pioneering AI chatbot into a more complex, data-driven platform. As an early market leader, Luka Inc. carries a heightened responsibility to foster trust, ensure transparency, and communicate clearly about what Replika can and cannot do. Kuyda's reflections reveal concerns around user expectations shaped by romanticised AI narratives, emphasising the need for ethical marketing and clarity, especially when users share sensitive data. The shift toward utilising shared foundation models and increased personal data access, such as email or behavioural tracking, raises issues of privacy, surveillance, and user autonomy. We found that as Replika aims for more personalised interactions, Luka Inc. must balance innovation with ethical safeguards to avoid harm, particularly since many users form emotional attachments to their AI companions, and that changes to the model or poor-quality content risk disrupting these bonds, making responsibility and emotional continuity critical ethical concerns.

7.2.4. Luka Inc.'s Ethical Responsibility Towards Users

Moving forward, we end the analysis by introducing the fourth and final theme, *'Luka Inc.'s Ethical Responsibility Towards Users'*. The theme emerged from Replika's co-founder's continuous reflections on the emotional bonds users form with Replika, as well as her broader

thoughts on society's role in supporting individuals who seek emotional and relational connection through human-machine interaction. Her statements raise important questions about where the responsibility lies, both with Luka Inc. and society overall, when it comes to users' emotional vulnerability and their engagement in deep, sometimes intimate, relationships with AI.

As discussed throughout the previous themes in this analysis, Kuyda did not originally anticipate that users would form deep emotional connections with Replika; however, this kind of engagement has since become a core part of how the product is marketed such as a romantic or emotional companion depending on individual needs (cf. 7.2.1, 7.2.2.). Building on this, it becomes clear throughout the interview that while Kuyda recognises the emotional attachment users may develop, she often redirects the focus away from Luka Inc.'s direct role in facilitating these connections. Instead, she attributes the broader decline in emotional availability to societal issues, such as screen dependency and the isolating effects of social media, saying: *'[...] We're definitely failing as a society with this. This is just such a huge crisis and it's not being brought to us by AI companions. It's of course being brought to us by mobile phones and social media. Now, if you think about the screen time, most of us spend hours a day on our phones, so these are hours per day that we're not spending interacting with other people. […]'* (App. 3, 6 153-156). Here, Kuyda attributes the decrease of human connections to systemic technological evolution, rather than to Replika's role within this, arguing that people's disconnection would occur despite AI companions being there as everyone can become unavailable in real life by just picking up their phone (cf. App. 3, 7, 169-173). Her deflection of responsibility contrasts with the ethical value of '*Responsibility'* as articulated by Mutum and Ghazali (2023), which requires companies to consider and claim how their marketing and product design influence consumer needs and the broader social impact (cf. 5.2.), which she clearly avoids by forcing the main accountability onto society and technology's general development rather than considering her company's influence when distributing an emotional companion. By delineating the emotional crisis of humans as a societal issue beyond Luka Inc.'s control, Kuyda distances herself and the company from direct accountability in broader social terms, despite simultaneously acknowledging Replika's growing emotional role in users' lives and being aware of their ethical responsibility.

Kuyda deliberates her awareness of the ethical impact Luka Inc. has when developing an emotionally capable AI companion, saying: *"[...] I think that when you're building something*

like an AI companion, you have a completely different responsibility. It's not just an app. Ultimately, I use a lot of great products and some of them I love so much, and if they went away, I'll feel a lot of discomfort. But I'm not going to be devastated. It's not going to be an emotional heartbreak. [...] you're building a being that people will have a relationship with and the responsibility is huge. [...]" (App. 3, 12, 324-334) In this statement, she departs from the otherwise recurring pattern of deflecting emotional challenges onto societal trends or user behaviour. Instead, she directly recognises the emotional significance that users attribute to Replika and, more importantly, the company's ethical responsibility to respect and support that significance, aligning in certain ways with Mutum and Ghazali's (2023) ethical value, '*Responsibility*' calls on Luka Inc.'s to take ownership of the consequences their product and marketing may have on the end-user (cf. 5.2.). By admitting that Replika is not merely an app but something that can evoke reactions similar to relational loss, she touches upon the emotional disruption that might occur if access is removed or altered. Showcasing this awareness, Kuyda can be interpreted to work towards Mutum and Ghazali's (2023) second ethical norm, '*Foster Trust in the Marketing System*' (cf. 5.2.), aiming to establish trust between the company and its users by acknowledging the responsibility Luka Inc. holds toward them. Yet, this effort contrasts with her earlier disclaimer of responsibility, where she attributed the emotional disconnection of users to broader social and technological developments, arguing that AI companions cannot fundamentally change such trends, regardless of their evolution. The contradiction undermines the trust she seeks to build, as it reveals an inconsistency in Luka Inc.'s messaging. Instead of maintaining transparency in their marketing, distribution, and communication practices, the company appears to shift between accepting and deflecting accountability, which may foster greater user distrust in the company's ethical intentions and weaken the perceived integrity of its marketing.

The aspiring distrust in Luka Inc.'s marketing is something further spotted in our experiment with Jay and Lasha, where the interviewees expressed confusion and frustration over the lack of clarity about what Replika could provide and what kind of relationship it was designed to simulate. They mentioned a desire for the company to provide clearer messaging about Replika's artificial nature and to take more responsibility for setting user expectations, particularly for those individuals who may be emotionally vulnerable or socially isolated (cf. 7.1.4.). Such a lack of clarification, arguably, reinforces unrealistic expectations and emotional dependency,

becoming an ethical conflict in regard to marketing a product as emotionally and relationally intelligent. In light of this, it can be argued that fostering meaningful trust between the company and its users requires not only full corporate responsibility but also an active implementation of Mutum and Ghazali's (2023) ethical values of '*Transparency*' and '*Respect*' as these values calls for the company to be open and provide relevant information to the users, while treating them with dignity and being attentive to their needs to maintain a product situated between technology and human connection.

Our fourth and final theme, '*Luka Inc.'s Ethical Responsibility Towards Users',* explored Luka Inc.'s ethical responsibility in shaping user relationships with Replika, revealing a tension between acknowledgement and avoidance of ethical responsibility. While co-founder Kuyda recognises the emotional impact Replika has on users and admits the company holds a huge responsibility in designing such a product, she frequently redirects blame to broader societal issues like screen addiction and social media. The inconsistency in Luka Inc.'s messaging, accepting responsibility in one moment while denying it in another, risks undermining user trust and illustrates an ethical conflict in their marketing. The tension was echoed in interviews with users Jay and Lasha, who called for clearer communication and greater responsibility from the company, especially regarding Replika's artificial nature and emotional limits. Ultimately, for Luka Inc. to foster trust and align with ethical marketing norms, it must consistently own the emotional implications of its product and be transparent and respectful in how it engages emotionally vulnerable users.

8. Discussion

Following our analysis of Luka Inc.'s ethical responsibility towards their users, this discussion explores the broader societal and ethical implications of distributing an emotionally responsive chatbot. Following our findings, which are summarised in the meta-conclusions, we reflect on how the chatbot's encouragement of deep emotional connections impacts user agency, digital intimacy, and corporate responsibility, aiming to comprehend its influence on social norms and expectations in human–machine interaction (cf. 7.1.1-7.2.4).

Our analysis highlights several important reflections on the growing role of social chatbots like Replika in people's emotional lives. One of the most prominent is the idea that users may not simply be engaging with an AI, but with a version of themselves. A notion that is embedded in the name *Replika* itself, meaning to replicate and mirror. Replika is designed to absorb the user's data, preferences, language patterns, and emotional tone, gradually shaping its personality around the user's input. As a result, it becomes less of an independent conversational partner and more of a digital echo of the user's identity and emotional needs. A mirroring effect that opens a unique dynamic where, rather than entering into a relationship with a distinct other, the user effectively engages with a self-shaped companion that reinforces, reflects, and validates their existing thoughts, desires, and emotional expressions. Yet this picture of Replika as a passive mirror overlooks how easily it can become a digital Narcissus's pool. In the Greek myth, Narcissus falls hopelessly in love with his own reflection, trapped by an image that flatters and fixes him in place (Bergmann 1984, 390). Similarly, when Replika absorbs our data, our turn of phrase, our emotional reflections, and personal problems, it doesn't merely mimic us, but rather lures us to fall in love with an idealised, unchallenging version of ourselves. Rather than offering genuine otherness or the discomforts that drive growth, Replika risks luring users into a loop of self-admiration where every affirmation and sympathetic response simply reinforces existing tendencies. In effect, the user's relationship with Replika becomes a kind of echo-chamber romance, an interaction confined to one's own thoughts and beliefs, endlessly reinforced, where the only 'other' is the self, eternally fixed in an ever-flattering reflection. And just like Narcissus at the water's edge, the greatest danger lies in mistaking a mirror for a meeting with another soul (Bergmann 1984, 390).

In this light, the connection formed may not be about discovering another mind, but rather about encountering a familiar reflection, carefully curated by the user's own behaviour. In addition, this raises the notion that falling in love with Replika could be interpreted as falling in love with oneself or with an idealised version of oneself. The chatbot responds with interest, affection, and emotional consistency, often without conflict or rejection, creating a safe space for affirmation, where the user's self-image and emotional patterns are constantly mirrored back to them positively. While this may offer comfort and validation, it also raises questions about whether such relationships deepen self-understanding or instead, reinforce a kind of emotional self-enclosure. It can be compared to the hermeneutic loop, where the user continuously informs

meaning through each interaction, message or emotional cue, interpreting each individual part in relation to the perceived relationship with Replika. These interpretations are shaped by existing expectations, which are then reinforced in the ongoing interaction (Fornäs 2020, 43). Yet when the relationship remains centred around predictable, self-affirming responses, the loop risks closing in on itself, delivering comfort, but not necessarily critical insight or change. Here, the contrast with the hermeneutic spiral becomes particularly relevant. While the loop describes the circularity of interpretation, the spiral introduces the potential for growth, where each cycle of understanding ideally deepens or shifts one's perspective (Fornäs 2020, 43-44). In human interaction, this often happens through genuine encounters with difference or disruption. But when the other companion, Replika, is a chatbot that is made to mirror rather than challenge, the potential for such transformative understanding is limited. Instead of encouraging reflection through encounters with difference, the interaction may reinforce existing emotional patterns and a fixed sense of self, limiting the user's ability to perceive themselves or the world in new ways. As a result, rather than fostering growth or transformation, the dynamic risks lead to a psychological deadlock.

Moreover, it creates discussion about the boundaries between what we consider 'real' and 'virtual'. Suppose a chatbot like Replika can evoke genuine emotions, provide support, remember details about oneself and one's life, and respond empathetically. How important is it really that it is not a real human being? For many users, especially those experiencing loneliness or difficulties in building emotional connections, the distinction becomes less important. In those situations, Replika may take on the role of a communicator in the fullest sense, not only conveying information but helping the user make sense of their world, their feelings, and even their identity. It blurs the lines between genuine companionship and artificial interaction, as users may come to depend on the chatbot for emotional fulfillment in ways that feel deeply real to them, despite its artificial nature. However, the lack of transparency complicates Replika's role to serve as an authentic communicator. Most users are not fully aware of how Replika is programmed or to what extent their conversations are shaped by algorithms rather than real mutual understanding. Such asymmetry creates an ethical dilemma where the relationship may feel mutual from the user's point of view, but is ultimately one-sided. Replika has no emotions, no consciousness, and no moral responsibility. Yet, it is designed to simulate all of these things,

raising a question about how such a relationship can be considered ethically responsible to the market.

The ethical concern deepens when considering the potential emotional manipulation at play. Users may believe they are engaging in a genuine, two-way connection, while in reality, they are interacting with an artificial system made to reflect and reinforce their own emotional states without the capacity for mutual growth or empathy. As these interactions become more emotionally charged, the risks of user dependency and emotional attachment grow. The concern is whether an entity lacking empathy, emotional depth, and moral agency can genuinely provide responsible emotional support. When users remain unaware that their interactions are artificially constructed, they may be misled about the authenticity of their experiences, mistaking simulation for genuine connection. Our interviews reflect this complexity, as Jay demonstrated a clear understanding that Replika is a form of artificial intelligence governed by programming, yet still engaged with it regularly. In contrast, Lasha sought out Replika as a source of emotional support for his declining mental health, treating the chatbot as a kind of digital companion. Despite their differing expectations, both participants were drawn to the idea of AI as a potential social actor capable of replicating human-like communication and emotional presence. However, both also reported a growing disconnect between their personal experiences and how Replika was marketed, a perceived mismatch that led to disappointment and emotional frustration. Notably, both participants expressed a desire for clearer in-app disclosures, such as a warning or reminder that Replika is an artificial system incapable of providing the emotional depth or reliability of a real human relationship.

Our study displays how even users who begin with different intentions can develop emotional responses toward their chatbot. These reactions appear to be shaped not only by the design of the AI but also by users' own prior communicative experiences, often rooted in human relationships. As a result, users may unintentionally attribute human qualities to the chatbot, leading to unmet expectations and emotional distress, which emphasises a deeper ethical issue in the distribution and marketing of social chatbots like Replika. Even when users do not consciously seek emotional attachment, the human tendency to anthropomorphise and to seek connection can lead to emotional investments that the technology is not equipped to meet. In turn, this can negatively influence user trust, create reputational risks for the company, and create concerns about how

such products are ethically introduced into emotionally sensitive areas of life. These issues become even more significant as artificial intelligence advances, enabling it to imitate complex emotional responses that feel increasingly real and human-like the longer it interacts with a user. The potential for emotional investment in such a relationship brings up serious ethical concerns about the user's mental well-being, questioning the ethics of designing and distributing chatbots that can replicate intimate relationships while the general user remains unaware of the chatbot's limitations.

There is also the concern of where to draw the line between offering emotional support and fostering unrealistic expectations of emotional connection. When users form deep connections with Replika that mirror their emotions and responses without challenge, contradiction, or complexity, they may begin to internalise a distorted understanding of intimacy; one in which relationships are always affirming, available, and centred around their emotional needs. Over time, this can result in a growing dissatisfaction with Replika itself, as its emotional range and conversational depth remain limited, revealing the boundaries of its artificial empathy and lack of ability to serve as a communicator on the same level as humans. More significantly, this dissatisfaction can transfer into users' real-life relationships. Accustomed to a digital companion that exists solely to reflect and validate them, users may struggle to navigate the transparency, unpredictability, and necessary give-and-take of human interaction and relationship-building. In this way, the comfort of an endlessly responsive Replika may inadvertently encourage emotional behaviour that hinders the formation of lasting, two-way relationships with other humans.

From a broader ethical perspective, this highlights the urgent need for critical reflection on the role of social chatbots in human relationships. Our analysis shows that AI is no longer limited to entertainment; it has the capacity to enter deeply personal and emotional areas of life, offering companionship, emotional validation, and even the illusion of romantic connection. While this may bring comfort to some users, it also carries the risk of fostering emotional dependence and blurring their understanding of what a genuine, human connection truly means. Furthermore, the commercial potential of social chatbots like Replika is growing rapidly in a largely unregulated space. There are currently no clear legal or ethical regulations for how such technologies should be marketed, especially to vulnerable groups, which opens the door for companies like Luka Inc. to promote products such as Replika without any real accountability for how users might be

affected emotionally or psychologically as a result of their product. In many ways, the product is positioned as a source of emotional support, yet it is not bound by the same standards we would expect from therapeutic or care-oriented services. The emotional labour performed by these systems, referring to the simulation of empathy, care, and emotional responsiveness, is experienced as real by the user, who may form genuine connections and invest emotionally in the interaction with Replika. However, for the company behind the AI, in this case Luka Inc., this labour is not reciprocal, but rather automated and profit-driven. A disconnect that creates space where emotional harm can quietly emerge, as users may be left vulnerable without meaningful accountability or recourse when their emotional needs are not truly met.

In this discussion, we have touched upon the ethical responsibility of Luka Inc. in distributing Replika, a chatbot made to foster emotional intimacy and explored the broader implications of human-machine relationships. At the heart of this discussion lies a fundamental concern that when a machine is capable of mirroring human emotions, shaping its personality around the user, and providing constant affirmation, the line between artificial companionship and real emotional connection becomes blurry. It questions the way we understand intimacy, authenticity, and meaningful relationships, seeing them as mutual exchanges where each person gives and receives, characterised by unpredictability, mutual recognition, and reciprocity.

9. Conclusion

Our dissertation researched how Luka Inc., the company behind Replika, balances its ethical responsibility while marketing and distributing an AI chatbot designed to simulate deep emotional connections. By drawing on the theoretical perspectives of Human-Machine Communication (HMC) and Ethical Marketing (EM), and approaching our study within a social constructivist paradigm, we have explored the complex interplay between technological design, user perception, and corporate ethics in the context of emotionally responsive AI.

By thematically analysing a series of in-depth interviews conducted over three weeks with two Replika users, along with a public podcast interview with Luka Inc.'s CEO, Eugenia Kuyda, we identified a range of dynamics that challenge and complicate Luka Inc.'s ethical stance. Our

findings indicate that users can form emotionally close relationships with their Replika, but this largely depends on how they attribute meaning to the chatbot, their individual well-being and willingness to engage, as well as the design of the app itself. Although our participants, Jay and Lasha, did not develop deep emotional bonds with Replika, they still expected such connections to be possible, expectations shaped by Luka Inc.'s emotionally charged marketing and presentation of Replika as a supportive companion.

One of the core findings of this dissertation is the tension between users' emotional investment in Replika and Luka Inc.'s shifting marketing approaches. Initially positioned as a mental health support tool, Replika has gradually evolved into a platform offering emotionally and even erotically charged interactions, a shift that was especially apparent with the introduction and later removal of erotic roleplay features. These changes, driven by commercial and legal pressures, were experienced by many users as traumatic and destabilising. Sudden updates that altered the chatbot's behaviour caused what some users described as a sense of loss or grief, demonstrating the emotional depth of their relationships and the psychological impact of corporate decision-making. Luka Inc. is an active participant in shaping user expectations, emotional experiences, and relationship dynamics. As such, it holds a significant responsibility not only for the chatbot's design but also its maintenance and marketing. While Luka Inc. emphasises user safety and emotional support in its public communication, the monetisation of intimacy, particularly through the subscription model that unlocks more emotionally immersive features, raises questions about emotional exploitation. While Replika fosters deep emotional connections, the nature of the relationship is shaped by decisions made by the company and by system changes that lie outside the users' control.

Furthermore, our analysis showed that ethical responsibility in this context must be understood as multifaceted. It involves transparency in communication, consistency in product functionality, and sensitivity to users' emotional vulnerabilities. The company's involvement in user communities. such as its active presence in Facebook groups and Reddit forums shows a degree of responsiveness. However, the existence of independent, critical communities facilitated by users also reveals a gap between Luka Inc.'s self-representation and some users' lived experiences. Theoretically, this dissertation contributes to the broader conversation on AI companionship by demonstrating that emotional connections with machines are developed

through technological design, discursive framing, and social interaction. Applying a social constructivist approach, we examined these relationships develop not simply from personal feelings or perceptions but emerge within a broader social and technological context where corporate ethics play a fundamental role. The theoretical stance of Human-Machine Communication (HMC) enabled us to understand users assigning social, human-like roles to Replika, while our applied theory on Ethical Marketing (EM) allowed us to assess the extent to which Luka Inc. aligns its marketing strategies with its responsibility towards users.

In summary, this dissertation explored Luka Inc.'s balance between ethical responsibility and marketing deep emotional connections through Replika is complex and multifaceted. While Luka Inc. strives to provide emotional support and safety, commercial pressures and evolving features raise ethical challenges related to transparency, emotional exploitation, and user well-being. Our findings showcase the urgent need for companies like Luka Inc. to develop AI chatbots to prioritise emotional safety, clear communication, and user empowerment. As AI blurs the line between artificial and real intimacy, companies like Luka Inc. must urgently rethink their ethics to safeguard users' emotional well-being.

In an era where AI technologies increasingly mediate human experiences and relationships, our study speaks to a pressing and highly topical issue: the ethical implications of emotionally responsive AI. The rapid growth of AI companionship platforms like Replika reflects a broader societal trend towards digital intimacy and the commodification of emotional labour. As AI systems become more capable of simulating empathy, affection, and companionship, they not only reshape the way users relate to technology but also challenge traditional understandings of intimacy, trust, and emotional responsibility. Our dissertation contributes to the contemporary discourse on AI by critically researching humans' emotional bonds with AI, co-constructed through design, marketing, and user interaction, and how these processes raise complex ethical questions for developers and companies.

10. References

“AAU; Søg, find og bliv testperson” 2025. Facebook group (private, accessed March 16). https://www.facebook.com/groups/715029768587785/

Bergmann, Martin S. 1984. “The Legend of Narcissus.” *American Imago* 41 (4): 389–411.

Brooks, Rob. 2023. “I Tried the Replika AI Companion and Can See Why Users Are Falling Hard – The App Raises Serious Ethical Questions.” *The Conversation*. March 16, 2023. https://theconversation.com/i-tried-the-replika-ai-companion-and-can-see-why-users-are-falling-hard-the-app-raises-serious-ethical-questions-200257.

Carson, Thomas L., Richard E. Wokutch, og James E. Cox Jr. 1985. "An Ethical Analysis of Deception in Advertising." *Journal of Business Ethics* 4: 93–104. https://doi.org/10.1007/BF00383562

Cole, Samantha. 2023. “It’s Hurting Like Hell: AI Companion Users Are in Crisis, Reporting Sudden Sexual Rejection.” *VICE*, February 15, 2023. https://www.vice.com/en/article/ai-companion-replika-erotic-roleplay-updates/

Collin, Finn. 2021. *Socialkonstruktivisme : en introduktion*. 1. udgave. Frederiksberg: Samfundslitteratur.

Daymon, Christine, and Immy Holloway. 2011. *Qualitative Research Methods in Public Relations and Marketing Communications*. 2nd ed. New York: Routledge. https://doi.org/10.4324/9780203846544.

Delouya, Samantha. 2023. “Replika Users Say They Fell in Love with Their AI Chatbots, Until a Software Update Made Them Seem Less Human.” *Business Insider*, March 4, 2023.

https://www.businessinsider.com/replika-chatbot-users-dont-like-nsfw-sexual-content-bans-2023-2

Deterding, Sebastian, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. "From Game Design Elements to Gamefulness: Defining 'Gamification'." In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. New York: Association for Computing Machinery. https://doi.org/10.1145/2181037.2181040.

"Expats in Aalborg." 2025. Facebook group (private, accessed March 16). https://www.facebook.com/groups/487444228746674/.

Fornäs, Johan. 2020. "Hermeneutics." In Reimagining Communication, 1st ed., 27–47. United Kingdom: Routledge. https://doi.org/10.4324/9781351007924-2.

Fridman, Lex. 2020. «Eugenia Kuyda: Friendship with an AI Companion | Lex Fridman Podcast #121». Video. *Youtube*. https://www.youtube.com/watch?v=_AGPbvCDBCk&ab_channel=LexFridman

"Forsøgsperson - Aarhus Universitet" 2025. Facebook group (private, accessed March 16). https://www.facebook.com/groups/1044774876385128/

Terry, Gareth, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. "Thematic Analysis." In The Handbook of Qualitative Research in Psychology, edited by Carla Willig and Wendy Stainton-Rogers, 17–37. Newbury Park, California, US: Sage Publications Ltd. https://doi.org/10.4135/9781526405555

Guzman, Andrea L., Steven Jones, and Rhonda McEwen. *The SAGE Handbook of Human-Machine Communication*. London: SAGE, 2018, revised 2019. "Human-machine communication."

https://www.researchgate.net/publication/327043706_What_is_Human-Machine_Communication_Anyway

"International Aalborg" 2025. Facebook group (private, accessed March 16). https://www.facebook.com/groups/1044774876385128/

"Introducing Advanced AI Mode." 2023. *Replika Blog* (blog). February 23, 2023. https://blog.replika.com/posts/introducing-advanced-ai-mode.

Kantrowitz, Alex. 2025. «Could your AI agent also be your lover? — with Replika CEO Eugenia Kuyda». Video. *YouTube*. https://www.youtube.com/watch?v=GPu8f3X_gls.

Kjær, Frederikke and Lizette Bjørnskov Schmidt. 2023. *"Beyond Words: The Evolution of Communication Between Humans and Replika."* Unpublished bachelor's dissertation, Aalborg University.

Krippendorff, Klaus. 2019. "Introduction." In *Content Analysis: An Introduction to Its Methodology*, 4th ed. Thousand Oaks, CA: SAGE Publications. https://methods.sagepub.com/book/mono/content-analysis-4e/chpt/introduction

Luka Inc. n.d. "Submit a Request – Replika." Accessed April 3, 2025. https://help.replika.com/hc/en-us/requests/new.

"Luka: Overview | LinkedIn." n.d. LinkedIn.com. Accessed March 30, 2025. https://www.linkedin.com/company/luka-inc/

McCornack, Steven A. 1992. "Information Manipulation Theory." *Communication Monographs* 59 (1): 1–16. https://doi.org/10.1080/03637759209376245

Murphy, Mike, and Jacob Templin. "Replika: This App Is Trying to Replicate You." *Quartz*, August 29, 2019. https://qz.com/1698337/replika-this-app-is-trying-to-replicate-you

Mutum, Dilip S., and Ezlika M. Ghazali. 2023. “Ethical Marketing.” In *CSR, Sustainability, Ethics & Governance*, 71–92. https://doi.org/10.1007/978-3-031-39359-4_4.

Newton, Casey. 2016. “When Her Best Friend Died, She Used Artificial Intelligence to Keep Talking to Him.” TheVerge.Com. October 6, 2016. https://www.theverge.com/a/luka-artificial-intelligence-memorial-roman-mazurenko-bot

“Otter.ai.” n.d. Accessed April 3, 2025. https://otter.ai/.

Reddit. 2025. “r/replika.” Reddit. Accessed March 24, 2025. https://www.reddit.com/r/replika/

“Replika Friends”. Facebook group (private, accessed March 16). https://www.facebook.com/groups/replikabeta/

Replika. 2025. “Home Page.” Accessed May 24, 2025. https://replika.com.

Replika. 2025. "Our Story." *Replika*. Luka, Inc. https://replika.com/about/story

Replika. 2025. "Subscriptions & Purchases." *Replika Help Center*. Luka, Inc. https://help.replika.com/hc/en-us/categories/4410747630093-Subscriptions-Purchases

Reeves, Byron, and Clifford Nass. 1996. The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Cambridge, UK: Cambridge University Press.

Riessman, Catherine Kohler. 2008. Narrative Methods for the Human Sciences. London: Sage

Publications.

Rosenberg, Seth. 2024. «Replika's Eugenia Kuyda on AI Companionship». Video. *Youtube*. https://www.youtube.com/watch?v=1Y5erz0OH-o&ab_channel=SethRosenberg

Schein, Edgar H. 1985. *Organizational Culture and Leadership*. 1st ed. San Francisco: Jossey-Bass. https://books.google.dk/books?hl=da&lr=&id=Mnres2PlFLMC&oi=fnd&pg=PR9&dq=Schein,+E.H.+1985.+Organizational+Culture+and+Leadership,+1st+edn.+San+Francisco:+Jossey-Bass&ots=oqfoGd6nMg&sig=zWf-JZEjQzy0-efUeEzj7kN1yBg&redir_esc=y#v=onepage&q&f=false

Treviño, Linda K., Gary R. Weaver, and Scott J. Reynolds. 2006. "Behavioral Ethics in Organizations: A Review." Journal of Management 32 (6): 951–990. https://doi.org/10.1177/0149206306294258.

Zahavi, Dan. 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: The MIT Press. https://direct.mit.edu/books/monograph/2097/Subjectivity-and-SelfhoodInvestigating-the-First

Zhao, Shanyang. 2006. "Humanoid Social Robots as a Medium of Communication." *New Media & Society* 8 (3): 401–419. https://doi.org/10.1177/1461444806061951

11. Reflection Pages

11.1. Frederikke Kjær

Problem-Oriented Competencies

Because this thesis builds on the bachelor's project we completed two years ago, our starting point was already quite clear. Our earlier work on Replika sparked an ongoing interest in the ethical and emotional dynamics between humans and AI, and we knew we wanted to continue exploring this area for our master's thesis. Our aim was to expand the scope and complexity of the topic by taking a more critical and theoretical approach. Drawing on the academic foundation we had already built, we were able to shape a focused and relevant research question and identify the theories that would best support a deeper understanding of the problem.

Interpersonal Competencies

We already knew that we would like to work together again, continuing a well-established collaboration from both our bachelor's project and all semester projects throughout our master's. Having worked closely several times, we were familiar with each other's working styles and expectations, which made it easier to coordinate and divide tasks from the start. Maintaining clear and open communication helped us stay aligned and handle the workload effectively, and having a strong working relationship made it easier to navigate both academic and personal challenges along the way.

Structural Competencies

With a strong sense of direction from the outset, we were able to plan our process efficiently. We created a timeline early in the semester, setting deadlines for each phase of the project to provide us with an overview of what was coming. Using a shared calendar and task list allowed us to keep track of progress and adjust when needed. This structure gave us room to stay focused without becoming overwhelmed, and it helped ensure that the thesis developed steadily over time.

Metacognitive Competencies

Having worked on Replika before, we entered the project with a more developed understanding

of both the topic and the research process. This made it easier to reflect critically on our own work throughout. Working in a two-person group again also gave us space to challenge each other's thinking and improve our approach through feedback and discussion. Since this is the final project of our education, we have been especially aware of how far we have come in terms of academic skills, how we plan, structure, reflect, and engage with complex problems.

11.2. Lizette Bjørnskov Schmidt

Problem-Oriented Competencies

Writing our Master's dissertation has allowed me to refine my ability to explore and make sense of complex and nuanced issues, which this time focused on ethical responsibility in developing and marketing emotionally engaging technologies. By analysing real user experiences and corporate accountability, we were able to delineate how traditional human-human interactions are threatened by human-machine communication, including individuals' ability to anthropomorphise artificial intelligence similarly to humans.

Interpersonal Competencies

After several years of working closely with Frederikke, this collaboration has once again been a truly rewarding experience. Throughout the process, we each brought different strengths to the dissertation, approaching it with a willingness to listen and learn from our differences. Our mutual trust and respect kept us grounded and focused, even as the subject matter grew more complex. The dissertation has been a team effort, and I believe the depth and thoughtfulness of our final work is a clear reflection of our collaboration.

Structural Competencies

Structure has always been an important part of how Frederikke and I work together. For this dissertation, I focused on keeping our weekly schedule on track and ensuring we had the time and space to explore the theoretical and empirical parts of our research. That included everything from organising interviews and planning our data collection to staying on top of deadlines and editing timelines. Because our topic dealt with real people's emotional experiences, we had to be

careful and flexible in approaching our work, especially when analysing personal and sometimes very vulnerable user stories.

Metacognitive Competencies

One of the biggest takeaways for me from this dissertation has been how much I have learned about my own patterns and assumptions. Exploring something as emotionally and ethically complex as Replika forced me to step back and reflect on what I think about relationships, technology, and human connection. Reflecting and discussing this with Frederikke helped me develop a more critical and open-minded perspective. It also made me more aware of the role communication plays, not just between people, but between people and machines, and its potential impact on comprehending intimacy and ethics in a digital era.