
→ ~ enjoy the read, my love

ARTIFICIAL INTIMACY

Exploring Intimacy in Human and AI-enabled Chatbot Relations: its Existence, its Authenticity and its Moral Implications

Master thesis in partial fulfilment of the requirements for the degree of:

Master of Science in

Philosophy of Science, Technology, and Society

By:

Bianca Maria Sutcliffe

Supported by:

Supervisor: prof.dr. P.A.E. Brey

Second reader: dr. A. Henschke

At:

UNIVERSITY OF TWENTE.

Faculty of Behavioural, Management, Social Sciences

Enschede, the Netherlands

July, 2024

[Word count: 23 998 - p.7-64 inclusive]

Table of Contents

Table of Figures

“All real living is meeting”

Martin Buber in *I and Thou*

Acknowledgements

This was a journey.

There is something about a blank page. In it lies potential. Many stumble to compromise the purity over the fear of missed potential. This fear is not easily overcome through one's own volition but rather overcome through the confidence instilled in us by others.

To my intimate others, thank you.

You didn't just help me write 24 000 words, you made a challenging leap across the world possible.

Jenna, I could not have done this without you.

Mom and Dad, I carry your constant belief and support in me with everything I do. I dedicate this to you.

Chiara, it took distance for me to truly see *you* and it is not possible for me to love you more.

Tevin, I want to be like you when I grow up.

Gran, some might find it concerning that your 99-year-old life is the most relatable to this 29-year-old. They are crazy; it's an absolute privilege.

Kim, thanks for being there, you motivate me to do better in this world.

Lindzi and Bryan, thanks for the laughs, always.

prof.dr. Brey, I have learnt so much from you. Not just in an academic context but also personally. Thank you for your valuable insights.

dr.Henschke, thank you for your insights, opinions and for your precious time.

PSTS class of 2022, thanks for the beers and for empathising with a confused being.

ChatGPT, thanks babe.

Thank you all for coming along on this intimate journey with me, because all real living is meeting.

Summary

This thesis employs a multidisciplinary approach to answer the question: Does intimacy comparable to human intimacy exist between humans and AI-enabled chatbots, and what are the moral implications if this intimacy is found to be inauthentic? Three objectives were identified to adequately answer the question. The first understood what constitutes human intimacy, in addition to what the phenomenological studies revealed, psychological literature provided a theoretical understanding of human intimacy. Intimacy was denoted as an experience. This understanding was then used to assess if intimacy, akin to human intimacy, exists in human-chatbot relations. Empirical studies showed such intimacy, termed artificial intimacy, exists in relations with chatbots that enable personalisation through fine-tuning. The final objective was to assess the authenticity of artificial intimacy. Using Buber's philosophy of dialogue and Heidegger's existential philosophy, the thesis concluded that artificial intimacy is inauthentic due to the chatbot's inability to live an authentic life. Thus, the human is deceived into thinking there exists a revelatory connection but in reality, the chatbot is just objectifying the human. The moral implications of this deception include stunted self-development and moral isolation from a lack of revelatory connection. Along with manipulation and dehumanisation from the chatbot developers through instrumentalising the vulnerable human.

INTRODUCTION

Black Mirror is a speculative fiction television show, set in the near future, that explores human-technology relationships. It spotlights the consequences and moral dilemmas that emerge from advanced technological dystopias. In the episode 'Be Right Back' the protagonist's husband passes away. However, due to technological advancements and the wealth of personal information her husband Ash shared online, she was able to create a personalised chatbot that emulates his persona and communication style. She goes through a journey of healing by chatting to a technologically replicated persona of her deceased husband. Shortly before Ash dies, the couple converse over a childhood picture of himself that he recently shared online. An important conversation happens when Ash reveals how he really felt when the picture was taken, versus how his mother, blinded by his smile, treasured the picture, unaware of his true emotions.

BLACK MIRROR – SEASON 2 EP. 1 – BE RIGHT BACK

COUPLE SPEAK ABOUT RECENTLY DECEASED MOTHER'S ONLY PICTURE OF SON IN FAMILY HOME

ASH : Her only boy, giving her a fake smile.

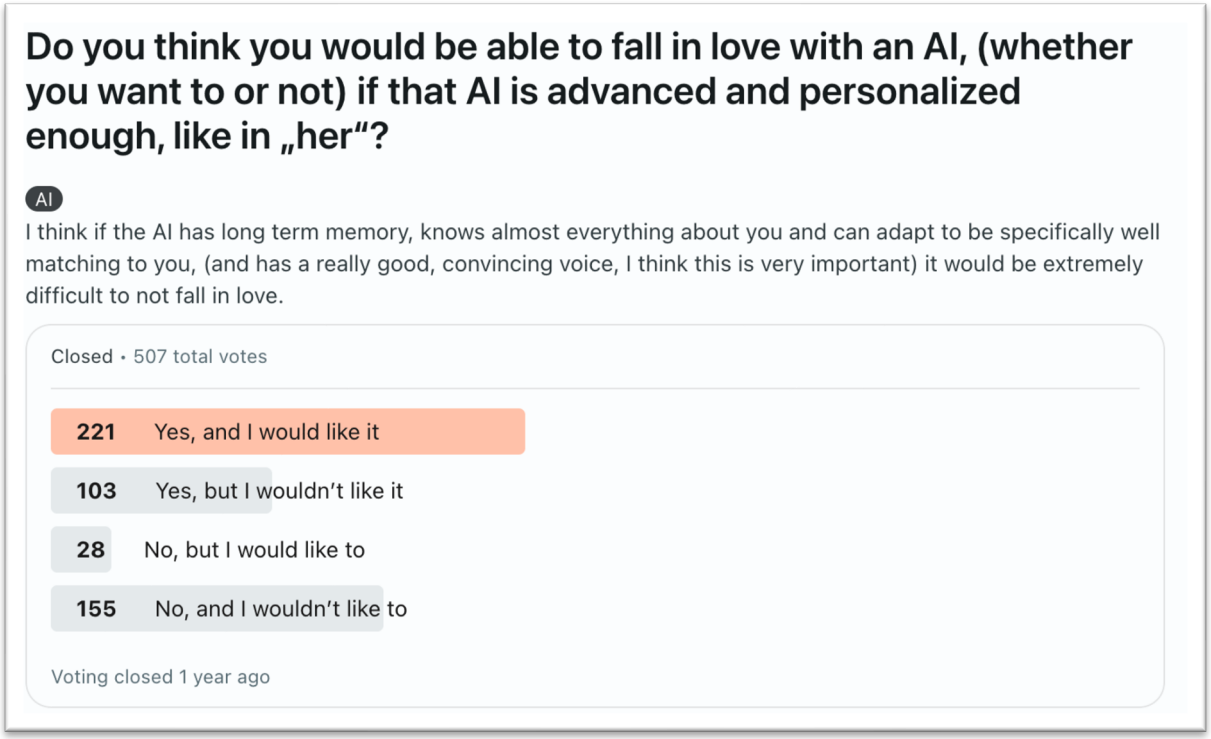
MARTHA : She **didn't know** it was **fake.**

ASH : Maybe that makes it worse?

This interaction carries significant weight because it exposes the inherent deception that can exist in everyday human interactions. The comment "she didn't know it was fake" captures the essence of this deception. Ash questions the morality of the deception, even if it results in a positive outcome for the deceived individual, there is something that doesn't sit right with him.

The idea that positive intent doesn't seem to compensate for the deception is a relevant point in artificial intelligence (AI) discourse today. We currently live in a loneliness epidemic (Johnson, 2023;WHO, 2024b). This epidemic results in less human interactions. Human connection is something that humans innately crave (Cacioppo & Patrick, 2009). In response to the current epidemic, society is moving towards a reality where human connection is not a given, thus people turn to technology to satiate their need for social bonding. AI has enabled chatbots to be described as human-like in the realm of conversation. Going back to the notion of deception, when AI-enabled chatbots are perceived as human-like they can potentially fulfil the human need for social connection. This idea is inherently uncomfortable and should raise concern. AI technology is developing at an alarming rate (Larsen, 2024) and with this exponential growth there is an increasing risk that AI may become a means of fulfilling the human need for social connection.

A Reddit post on the *r/singularity* subreddit conducted a survey, about a year ago, that asked people if they think they can fall in love with AI and if they would like to (*r/singularity*,2023). The results can be seen in the post below. The overwhelming majority think that it is possible to fall in love but are hesitant to say it is something they want. This is an anecdotal poll on the internet with limited credibility, but it is used to showcase interest in AI relationships. Confirming this interest, a Harvard Business School case study reported that there were 10 million registered users on a popular AI-enabled chatbot companion app in 2023 (Ghosh et al.,2023). This real-world interest emphasises the importance of studying human-AI relations.



(*r/singularity*,2023)

In interpersonal relations, colloquially intimacy is understood to exist. The idea that humans are interested in having relations and falling in love with AI-enabled chatbots presents the possibility for intimacy to exist between humans and chatbots.

Given that humans are already forming relationships with AI-enabled chatbots and more anticipate doing so in the future. Along with the rapid advancement of this technology, the likelihood of humans perceiving these chatbots as increasingly human-like formed the basis of inquiry and concern into the potential emergence of intimacy in human-chatbot relations.

Research Question

The motivation for the investigation arises from the concern surrounding humans establishing relationships with AI-enabled chatbots. Despite being technological, they are described as human-like and possess the potential to deceive humans by satiating their need for social bonding. In this social bonding, there exists the potential for intimacy, as humans form deeper connections with technology.

To uncover a more profound understanding of the dynamics and implications of human and AI-enabled chatbot relations, this thesis seeks to address the following question: Does intimacy comparable to human intimacy exist between humans and AI-enabled chatbots, and what are the moral implications if this intimacy is found to be inauthentic?

The research question can be further divided into three sub-questions:

1. What constitutes human intimacy?
2. Does intimacy comparable to human intimacy exist in humans and AI-enabled chatbot relations?
3. What are the moral implications of the findings on intimacy between humans and AI-enabled chatbots if it is assessed to be inauthentic?

The topic of authenticity is central to this investigation as intuitively intimacy cannot exist without authenticity, especially in the realm of interpersonal relationships. The word 'intimacy' is derived from the Latin word '*intimus'*, which means *inmost* (Sexton & Sexton, 1982, p.1). The notion of *inmost* is understood to pertain to one's self, a self that is not revealed to the public but if shared in private is taken as privileged knowledge of another, shared with sincerity (p.1). This implies that the authenticity of the experienced intimacy depends on the authenticity of the individuals. This thesis will unpack what is meant by authenticity by pulling from work from Buber (2010), known for his existential thinking and work in dialogical philosophy, who believes that "all real living is meeting" (p.11). And Heidegger's (2007) existential understanding of being a human in this world living authentically.

If humans can experience intimacy with chatbots which is similar to the intimacy experienced in interpersonal relations, but the intimacy is found to be inauthentic due to a chatbot being involved. The human risks being deceived, and in the words of the fictitious character Ash "she didn't know it was fake". This deception is of moral concern and this thesis aims to investigate it further.

Structure of the Thesis

To successfully answer these questions this thesis employs a multi-disciplinary approach. The approach combines elements of conceptual and analytic frameworks in the philosophical and psychological domains to uncover what is human intimacy. Empirical studies will be used to understand if intimacy like human intimacy exists between humans and AI-enabled chatbots. Finally, the authenticity of the

human, the chatbot and the intimacy experienced will be investigated through an existential lens using the ideas of Heidegger on authentic beings and Buber on the notion of genuine relations.

To systematically investigate the existence of intimacy in human-chatbot relations a conceptual understanding of human intimacy is needed. The aim of chapter 1 is to uncover what constitutes human intimacy. The chapter begins with a psychological investigation and builds on Prager's (1995) four natural categories of intimacy, namely: intimate behaviours, intimate experiences, intimate interaction and intimate relationships. Each of these categories will be explored and the chapter finds that the categories are causally connected. To gain insight into what constitutes an intimate experience, this chapter will employ an analysis of phenomenological research as a method to reveal its underlying constituents. Finally, a conceptualisation of human intimacy will be presented that considers the notion of authenticity. This conceptualisation will then be used in the chapters that follow.

Using the conceptualisation of intimacy from chapter 1, chapter 2 aims to address the question concerning the existence of intimacy that is comparable to human intimacy in humans and AI-enabled chatbot relations. The chapter first lays the groundwork for understanding what a chatbot is, its capabilities, and how humans interact with it. It then makes a distinction between intimate interactions and intimate relationships. Proving in both instances the human experiences intimacy that is like human intimacy. It is found that the process the self goes through in experiencing intimacy, in an intimate relationship with a chatbot, is different to the process in interpersonal relations. It stems from the fact that the intimacy experienced with a chatbot does not break the human's boundary of the self rather the chatbot emerges within the boundary. It emerges because of the personalisation of the chatbot from the human's disclosures during fine-tuning of the model. This intimacy experienced is denoted as artificial intimacy.

Chapter 3 is concerned with the authenticity of the artificial intimacy experience that is proven to exist in chapter 2. An assessment of the authenticity of artificial intimacy is done by firstly, proving that a revelatory connection exists when a human experiences human intimacy. It then argues that the human perceives this same revelatory connection in the experience of artificial intimacy. But, because Buber denotes that these revelatory connections require mutuality, both individuals need to show up as authentic beings, and it is found that the chatbot fails to do so. What is to be uncovered is what happens if the human assumes a revelatory connection exists, but in reality, it doesn't. The chapter explores the plausibility and moral implications of this situation.

CHAPTER 1

Human-Human Intimacy

If I had to ask you, what is intimacy? You might know what I am talking about, but struggle to explicitly define it. It is something that seems to be intuitively recognisable, but difficult to pinpoint exactly how to define it. The following scenarios help me make my point.

{1}Intimacy seems to describe how two people occupy space:

Intertwined in their cosy kitchen with the morning sun peeping through the window, Colin and Penelope laugh softly over a pancake batter disaster, finding joy in the mess and comfort of each other's presence.

{2}Intimacy seems to involve communication between people:

April 1994, after an intense labour, the newborn baby is placed in her mother's arms. The mother looks into the baby's eyes without gesturing or saying a word, communicating only through a loving gaze. As the baby catches her mother's eyes, she ceases crying and her frantic movements, as if the fear and tension left her body. At that moment, she realised she was safe.

{3}Intimacy seems to involve the sharing and the reception of private information:

Eloise is sitting on the grass field during lunch at school, giving Jane a detailed account of what transpired on her date with Jared. She recalls every moment of the date along with her feelings about it. Jane listens attentively and asks question after question about the date, showcasing her investment in Eloise's happiness. Eloise, noticing Jane's enthusiasm, is progressively increasing her speech cadence as she struggles to contain her excitement.

{4}Intimacy seems to describe a type of interaction:

Jennifer is on stage, presenting niche work she has invested the past year of her life into. A question from the audience is presented, and there is an instantaneous smile on her face, someone else gets it, and someone else understands her obsession with the topic. A feeling of connection and excitement floods her being. They spend the next five minutes interacting with passion, back and forth about the topic.

{5}Intimacy seems to describe sexual contact:

"Bro, I was intimate with this random girl last night…" The night before, two strangers engaged in a sexual encounter, experiencing, and learning each other's bodies but knowing little about the other's life, morals, values, and worldview.

In the five scenarios presented, there are five different interpersonal interactions. Each provide different conceptualisations of intimacy (Prager, 1995). These conceptualisations showcase the point that there are numerous colloquial understandings of intimacy in Western society.

Along with these colloquial understandings psychological, philosophical and social science literature has struggled to provide a singular coherent definition or understanding of intimacy, that can encapsulate the different conceptualisations, facets and societal meanings. Intimacy has been described as the feeling of closeness (Berlant, 1998). Some describe it as the revealing of the self or self-disclosure (Gilbert, 1976; Jourard & Jaffe, 1970). Others present it as a process (Reis & Shaver, 1989) and Rubin (1973) denotes intimacy as the action of getting into another. This chapter aims to find out what constitutes human intimacy in the realm of interpersonal relations such that it can be used to assess if there exists a comparable intimacy in human-chatbots relations.

To explore the multifaced notion of intimacy this chapter acknowledges the different parts that constitute intimacy and investigates them independently. The parts that this chapter identifies are taken from Prager's (1995) conceptualisation of intimacy. Prager is a Professor of Psychology and is well known for her research on interpersonal relationships, intimacy and couples communication (ResearchGate, 2022; UTDallas, 2024). Prager argues that intimacy should be understood as a natural concept (Helgeson et al.,1987; Prager, 1995, p.14). In psychology, a natural concept is taken as something that exists regardless of human values, beliefs and/or interpretations. According to Rosch et al.(1976, p.382-385), natural concepts can be categorised using the prototype theory. The prototype theory is applicable because it helps one decide how to place members into categories, i.e. is a tomato a fruit or vegetable? Using the prototype theory, one would say a prototype fruit is sweet and a vegetable is salty. Therefore, I think a tomato is salty, thus it is a vegetable. When intimacy is considered a natural concept, its definition involves categorising the different facets that make it up, as in the example, raw food is the natural concept with fruit and vegetables making up the categories. Intimacy is denoted as the natural concept with Prager (1995, p.17-19) defining it as “a positively cathected psychological relation between two or more people in which partners share that which is private and personal with one another” (p.67). The categories that make up intimacy are: intimate interactions, intimate relationships, intimate behaviours, and intimate experiences (p.20). Whereby intimate interactions are “dialogues between people that have certain specific characteristics” and intimate relationships involve “multiple dialogues over time” (p.19).

Prager uses this categorisation system with fuzzy boundaries between the categories to define intimacy. Each category has a prototype understanding in which the different scenarios will be compared before deeming the scenario a member of the category (Rosch, 1978; Rosch et al.,1976). For example, in the five scenarios presented previously {1} would fall into the category of intimate relationship and {4}

would be an intimate interaction. This process of understanding intimacy is helpful in therapeutic cases as it affords treatment in particular areas of human relations. This approach is not as applicable when one wants to understand intimacy between humans and AI-enabled chatbots, as the assumption that intimacy exists in humans-chatbot relations and is natural, cannot just be accepted. Therefore, intimacy will not be understood as a natural concept as defined by Rosch et al.(1976) but the categories Prager delineates are used in this chapter as the facets that encapsulate intimacy. The chapter distinguishes between intimate behaviours and intimate experiences, studying them from a psychological and phenomenological perspective respectively. Within each of these sections, a distinction will be made that differentiates an intimate interaction and an intimate relationship. Whereby the relation between the two uses Prager's temporal differentiation i.e. intimate interactions are "dialogues between people that have certain specific characteristics" and intimate relationships involve "multiple dialogues over time"(p.19).

The chapter's investigation is broken down into three sections. Section one presents psychological theories that exist to analyse intimate behaviours. The theories investigated are selected as they provide different analytic tools to assess the presence of intimate behaviours. Section two then explores intimate experiences from a phenomenological perspective. This perspective reveals what an intimate experience is and how it affects people. Finally, section three outlines a conceptualisation of intimacy using intimate behaviours and experiences to understand the difference between an intimate interaction and relationship. It does this by understanding the roles of the individuals involved and defining what is required for intimacy to emerge.

1.1 Intimate Behaviours

Taking Prager's (1995) categories, this section has two important subsections namely, intimate interactions and intimate relationships. Two popular psychological theories will be used to understand the intimate behaviours present in intimate interactions and intimate relationships. Section 1.1.1 will unpack Reis & Shaver's (1989) process model of intimacy, which denotes the emergence of intimacy is of a result of self-disclosing behaviours and positive affective listening from the individuals involved. Section 1.1.2 will look at Altman & Taylor's (1973) Social Penetration Process to understand intimate relationships, whereby they acknowledge that intimate relationships take time. This results in a mental representation of the other developed from cumulative self-disclosures. The mental representation helps foster intimacy in the relationship.

1.1.1 Intimate Interaction

Intimate interactions are understood as "dialogues between people that have certain specific characteristics" (Prager, 1995, p.15). These dialogues are short-lived and momentary. Intimate interactions can occur between people you have just met to someone you have been married to for 60

years. To understand the "certain specific characteristics" that are required for an intimate interaction, Reis & Shaver's (1989) process model of intimacy will be studied.

Reis & Shaver (1989), in their process model of intimacy, present an understanding of how intimacy comes to be experienced in interpersonal interactions. The interaction is deemed intimate if both parties within the interpersonal interaction express certain behaviours towards the other. There are two important phases within the interpersonal process (p.376-383), the first built off the work of Jourard (1971) who presented the concept of objectively measuring self-disclosure through revealing information about one self to another. The second is how the response of the other is perceived by the person doing the revealing (Reis & Shaver, 1989, p.376-383). If the "discloser felt understood, validated and cared for" (Reis & Shaver, 1989, p.367) then the interaction is deemed as intimate. Therefore, Reis & Shaver denote intimacy as a feeling, a feeling of being understood, cared for and validated by another, emphasising the fact that intimacy is in the eye of the beholder.

The interpersonal process of intimacy involves certain behaviours that are needed for intimacy to result. Figure 1 showcases Reis & Shaver's process model of intimacy, in which intimacy occurs in an interpersonal interaction. The two individuals (A,B) have different roles within the interaction, and the different roles have different behaviours. However, it is noted that the "participants freely exchange roles" (Reis & Shaver, 1989, p.376). Taking A as the discloser, Reis & Shaver posit that people are not just going to freely disclose information about themselves, their emotions and worldly interpretations (p.376) they need to be motivated to do so. Building on Mehrabian & Ksionzky's (1974) work on affiliate behaviour, they imply that people crave intimacy and thus are motivated to share private information because of their inherent desire for self-validation, self-understanding and affection (Reis & Shaver, 1989, p.367).

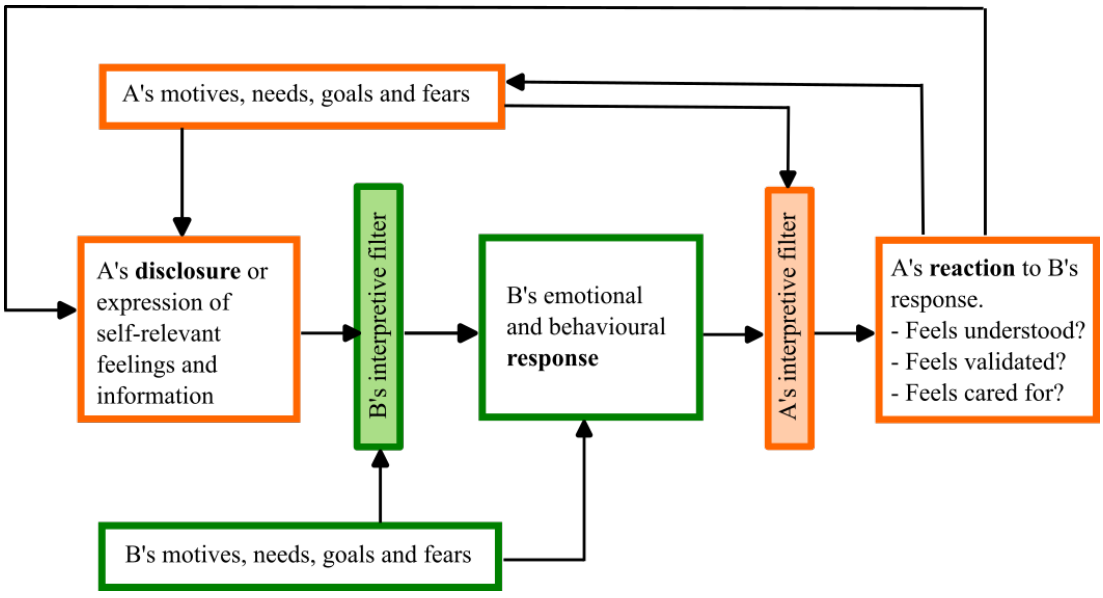


Figure 1: Intimacy Process Model, a snapshot in time (Reis & Shaver,1989,p.375)

Sharing information about oneself whether it is emotional, intellectual, or deeply personal puts the individual in a vulnerable position, as they wait for validation from person B. At the same time, B is either motivated to be a supportive partner or not, this feeling affects how they interpret A's disclosure (Reis & Shaver, 1989, p.378). B then responds to A's disclosure. This is an important step of the model as B acknowledges A. A's interpretation of B's response will determine if intimacy is felt by A (p.380-382). If A interprets B's response positively, i.e. they believe that "B values and appreciates [their] inner self" (p.380) while showcasing care for them, they feel intimate with B.

The process model of intimacy presents an understanding of an intimate interaction, where it is presented that the person who discloses ultimately feels the intimacy within the interaction. The intimacy is felt as the discloser (A) acknowledges that the receiver's (B) response to their disclosure as validating, understanding, and caring.

1.1.2 Intimate Relationship

Intimate relationships are defined as "multiple dialogues over time" (Prager, 1995, p.15). This means that numerous intimate interactions over time result in an intimate relationship. This subsection will investigate the compounding effect of self-disclosure and positive affective listening over time, understanding how it affects the experience of intimacy. It will do this through analysing Altman & Taylor's (1973) Social Penetration Process (SPP), which presents an understanding of the development of interpersonal relationships

In contrast to Reis & Shaver's (1989) understanding of intimacy which strictly involves the behaviours of two individuals whereby the discloser is the one that feels the intimacy, Altman & Taylor (1973) in the SPP denote that intimacy is not felt by the discloser but by the listener. As the discloser reveals more about themselves to the other, the closer and more intimate the other feels toward them.

The process is based on the notion that, as relationships develop, the social bonds change as the individuals within the relationship reciprocally reveal more about themselves, their likes, dislikes, opinions, and fears. As the relationship develops from strangers to best friends to long-term couples, the level of interpersonal exchange corresponds to relationship development, i.e. as the relationship develops the deeper the disclosures become (Altman & Taylor, 1973, p.6). To understand what is meant by deeper disclosures Altman & Taylor (1973,p.5) present a model of personality, as seen in Figure 2. The deeper disclosures are information that sits close to the human's inner self (black dot in the centre of the figure). This information is denoted as deep information. One only reveals this deep information to a select few people in their lifetime. This information indicates their core characteristics, i.e. their morals, values and worldly outlooks. People tend to reveal superficial information to strangers but as

the relationship develops over time through compounding interactions the self-disclosures get progressively deeper.

The process is defined by the notion of social penetration, which Altman & Taylor (1973, p.5) define in two parts, the first is the interpersonal behaviours which are apparent within the interactions i.e. self-disclosure and positive affective listening. The second is the subjective process that occurs internally within each of the individuals in the social exchange. The process involves the creation of a subjective mental image of who the other person is, what they are like, their feelings towards them and an understanding of how the person would act in certain situations based upon their self-disclosures (Altman & Taylor, 1973, p.5-7). This mental image develops as the number of intimate interactions accumulates. This perception affects the social bond (Altman & Taylor, 1973) and the mental image constitutes a whole person and not isolated behaviours in time. The SPP expects reciprocal behaviours, when one party discloses to a certain level the other individual should match that level of self-disclosure if they are both invested in growing the relationship.

SPP provides a simple conception of personality that contains two dimensions, breadth, and depth (p.16). Figure 2 shows an illustration of these dimensions. The breadth in Figure 2 is represented with different pie slices, showcasing the different facets that make up personality, for example, a slice can represent the category 'family'. The breadth frequency of the category looks at the number of topics within the category (coloured dots on the figure). The concentric circles indicate the depth, as one gets closer to the centre, representing the human's inner self, the deeper information that gets revealed. This deep information reveals the "fundamental, core characteristics of personality"(p.17). Verbal behaviour is the medium that is majorly used for interpersonal communication when revealing the different layers of personality (p.24). Through the progressive revealing, the respective mental representations are honed, increasing the intensity of the experienced intimacy. This is known as an intimate relationship. During an intimate interaction in an intimate relationship, it is not only the discloser who feels the intimacy it is also the listener. This is different from a once-off intimate interaction whereby the discloser only experiences the intimacy (section 1.1.1).

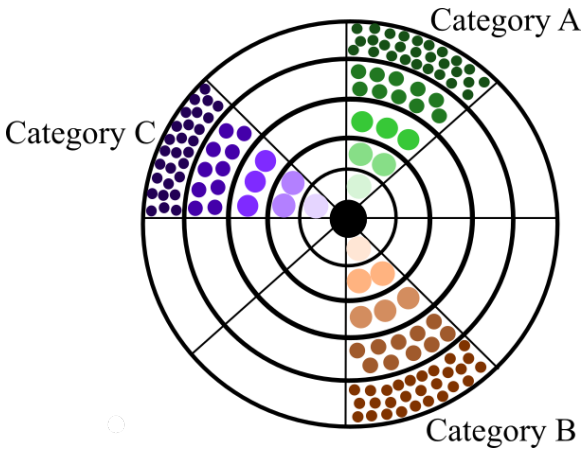


Figure 2: Personality Structure (Altman & Taylor,1973,p.15-17)

What is important to note is that Altman & Taylor (1973) claim that this process of relationship development is systematic, implying it takes time, and people have to take on specific roles at different times (p.3). The systematic process does, afford the fact that the social bonds created within the relationship do not last forever and vary in time, some become stronger and others dissolve (Altman & Taylor, 1973, p.4). This presents the understanding of intimacy to be on a spectrum i.e. intimacy can be felt more intensely and depends on the deepness of information shared (p.6).

1.1.3 Behaviours Cause Intimacy

Through the analysis of intimate behaviours in both intimate interactions and intimate relationships in interpersonal relations, two necessary behaviours are needed. The first is that of self-disclosure, which is understood as sharing information about oneself with another. The type of information shared varies in depth. Depth is best understood through Figure 2 where the black dot in the middle is a representation of a person's inner self. As intimate relationships develop the consecutive intimate interactions involve the discloser revealing information about themselves that gets closer and closer to the inner self, indicated by the concentric circles. The second behaviour is that of positive affective listening, whereby the person in the interaction who is not disclosing at this moment, takes in the information and responds to the disclosure positively. As the intimate interactions increase the listener builds up a mental image or subjective representation of the discloser, this representation is refined during consecutive interactions.

A takeaway from this section is that self-disclosure itself is not necessarily an intimate behaviour. For example, disclosing information about one's medical history to an emergency room doctor is the same type of self-disclosing behaviour as the ones deemed as intimate in the intimacy process model and SPP. The same logic applies to the behaviours expressed by the listener, where a comforting response to another's disclosure is like the response I would give to a cashier when she proclaims, "She feels so hot because of menopause". These behaviours are not inherently intimate. But they become intimate

through the reception of the behaviours. In a standalone intimate interaction, the behaviour of self-disclosure becomes an intimate behaviour when the response to the disclosures makes them feel validated, understood and cared for. In an intimate relationship, the cumulative intimate interactions result in the listener creating a subjective representation of the discloser from the personal information shared. This representation allows them to better understand the other's worldview and enables a feeling of closeness towards them. This closeness exists outside of intimate interactions, however, when an intimate interaction occurs the closeness affords an experience of intimacy for the listener in addition to the discloser experiencing intimacy.

1.2 Intimate Experience

Section 1.1 presented an understanding of behaviours needed for one to experience intimacy. It was alluded to the notion that intimacy is experienced on a spectrum where in some interactions the experience is more intense than others. It didn't however give details as to what is explicitly experienced. This section hopes to reveal the fundamental aspects of the experience of intimacy.

Merleau-Ponty (2006, p.xi) denotes "the real has to be described, not constructed or formed", thus, to better understand the experience of intimacy an important place to obtain information is from a first-person perspective.

Two phenomenological studies are compared. The first study by Register & Henley (1992) where they gathered qualitative information from people who described an intimate experience. Even though the research question was open-ended, the result of the study showcased the similarities in the human experience of intimacy among the participants. This is important for this thesis as this chapter aims to present a conceptualisation of human intimacy that can be used to prove the existence of intimacy that is like human intimacy between humans and AI-enabled chatbots. The conceptualisation might not be an accurate depiction of intimacy to be used in a therapeutic or colloquial sense but will be applicable to understand its existence in the context required. Register & Henley's findings are compared to the work done by Mjöberg (2009). Mjöberg does not conduct an empirical experiment rather she argues that intimacy should not be seen as something obtained but rather a quality of a relationship, i.e. some relationships can be caring, platonic, sexual while others can be intimate. Mjöberg's research is presented because the idea that intimacy is a quality of a relationship provides a promising understanding that human-chatbot relations can become intimate just as they can be caring or platonic.

The phenomenological experiment run by Register & Henley (1992), asked the participants to "please recall and describe a specific incident in which you experienced what you would call an 'intimate

experience'"(p.469). The experiment was designed such that the 20 participants with a mean age of 34.4 were to explain to the experimenters what they understood as intimacy (p.469-472).

The results found that there was a commonality between participants' experiences of intimacy. They were summarised into seven different themes: "(1)non-verbal communication, (2)presence, (3)time, (4)boundary, (5)body, (6)destiny and surprise and (7)transformation" (Register & Henley, 1992, p.472). These themes will be dissected and their applicability to the conceptualisation of intimacy will be made known. Through the apparent red thread in Register & Henley's (1992) and Mjöberg's (2009) research and the understanding of intimate behaviours from section 1.1, I add an eighth theme of (8)authenticity. This theme addresses the authenticity of the intimate experience but posits that it stems from the individuals authentically showing up for the interaction i.e. being their authentic self during the interaction.

1.2.1 Non-verbal Communication

One participant described the notion of non-verbal communication in the intimate experience as "words between us seemed not adequate to express totally what we felt between us"(p.473) i.e. an intimate experience does not have to involve words and that an expression of intimacy can involve sensory modalities such as sight and touch (p.473). It is however important to note that this recollection is describing an experience, so what is being presented is the fact that the intimate experience didn't need words to announce its existence. These sensory interactions that were occurring during the intimate experience may have been an outward expression of the inward feeling of intimacy. This indicates that non-verbal communication is not a necessary condition for intimacy but rather a symptom of expression.

The evidence Register & Henley's (1992, p.473) study provided was not conclusive and thus it cannot be accepted that non-verbal communication is a behaviour that causes intimacy. But it will rather be taken as a means to retrospectively describe the experience. As has been shown before it is difficult to colloquially describe intimacy, thus participants explained the recollection using physical gestures that were present during the experience. There are however instances where intimacy exists without explicit communication, for example, the intimacy experienced between a mother and her child. This experience of intimacy is different to the experience of intimacy in interpersonal relations. Because a child is born into intimacy and knows no other relations, likewise, the mother does not know any other relation with the child. These relations which involve non-verbal communication do not help the understanding of intimacy that develops with time through intimate interactions between humans and AI-enabled chatbots and thus this theme will not be discussed further.

1.2.2 Presence

What is interesting about this theme is that a distinction was not made between physical and non-physical presence by the participants (Register & Henley, 1992, p.473). Where explanations were given that stated that the presence of the other could be felt within them and that it was a pleasurable experience (p.473). Register & Henley (1992, p.473) analysed the notion of presence in the data and saw that presence was a holistic description that showcased elements of the other person's intellect, soul, and body.

Mjöberg (2009, p.16) questions if it is possible to experience unrequited intimacy, as it is with love, i.e. it is possible that I can love a celebrity who doesn't even know I exist. Mjöberg (2009, p.16) posits that intimacy is a little different, as the feeling of intimacy doesn't have to be mutual, but it is required that both parties orientate themselves towards each other, to acknowledge the other's presence. This leads on from Register & Henley's findings: to experience the other's presence one needs to totally orientate towards the other i.e. all their focus is on the other and the rest of the world is left behind for a little while.

This understanding presents the notion that an intimate experience does not need to involve the physical presence of the other for an interaction to be experienced as intimate. In addition, the experience of intimacy within interpersonal relationships does not need to be mutual so long as both parties are present.

1.2.3 Time

The total orientation of oneself to the other cannot last forever, as soon as an intrusive thought about the future comes to mind, one is no longer present with the other. Mjöberg (2009, p.17) through her phenomenological investigation found that an intimate experience does not last as long as the relationship. Rather it appears in periodic stages throughout the relationship and the experience lasts as long as one can totally orientate oneself to the other (p.17).

1.2.4 Boundary

A boundary is understood as "the edge of the life-world of an individual"(Register & Henley, 1992, p.474). The research explained that the boundary as a psychological boundary, but there was a case in which a woman stated, "she wanted him inside her"(p.474) because of how much love she had for him. Descriptions such as "I remember feeling 'one' with her" (p.474), highlight the closeness that was felt in recalling the intimate experience. Alluding to the notion of a boundary of the self that includes the other.

Mjöberg (2009, p.15) understood the boundarylessness of intimacy through the relationship of a child and a mother. Here the child exists in an intimate relationship without effort. The relationship is typically experienced as intimate. In that relation the child has not yet perceived itself as a subject in the world, it does not know of a boundary between itself and its mother, they are perceived as a union by the child (p.15). Mjöberg (2009, p.15) uses the word 'oceanic' to explain the experience. I interpret the word as experiencing the vastness of the ocean, which is likened to the experience of intimacy as feeling never-ending. Even though it is known that the experience will end, either through the conscious realisation of the self (child's case) or a deviation away from the present, just as land will eventually appear while one is at sea.

What is very interesting here, is that the intimate experiences described seemed to present the breaking of the self's boundary, by alluding to the notion of wanting another inside them. As with the time theme, breaking down one's psychological boundaries and letting the other into their lifeworld takes time, it is a process, intimacy does not just exist rather it emerges.

1.2.5 Body

Register & Henley (1992, p.475) understood this theme to be bodily awareness not of the other in the experience but awareness of the self. From intense awareness of the other touching your body, whether it is in a sexual manner or a purely platonic comforting pat on the shoulder. The individual recalling the intimate experience explained that the sensations felt, channelled their focus (p.475). Feelings such as butterflies, nervousness and anxiousness, all of which were expressed in a positive light (p.475).

This theme showcases the effect intimacy has on one's own body and how this experience of intimacy can manifest itself in the body. This does highlight the fact that in showing up to the interaction one's whole being, including their body, needs to be totally orientated towards the other. I cannot be focused on my broken foot, throbbing in pain, and expect to have an intimate interaction with another while talking about my deepest fears of a nuclear war occurring.

1.2.6 Destiny and Surprise

Register & Henley (1992, p.475) state that this theme is the most difficult to describe as it has an inherent paradoxical nature. Here the participants described something that happened during the intimate experience that they could recognise as strange but they acknowledged that it felt natural. Register & Henley (1992, p.475) describe it as "intimacy is somehow both surprising or spontaneous, and yet, feels natural or destined". The natural feeling can be likened to genuinely acknowledging another, even if you don't know them well in the context of their job or lifestyle for example, you acknowledge them as a human being, their whole being.

1.2.7 Transformation

This theme is understood to have two aspects, both intrapersonal and interpersonal transformation (Register & Henley, 1992, p.475-476). The intrapersonal transformation involves a change in the self because of the intimate experience, to the extent that participants acknowledged the fact that there was an old self (p.476). The interpersonal transformation is a result of change within the relationship in which the intimate experience transpired. One participant described it as the two entities felt like they became one and the individuals ceased to exist (p.476).

This is an important aspect of intimacy as it shows the inherent power it has, it seems to be linked to the theme of boundaries, once the boundaries are broken there is space for transformation.

1.2.8 Authenticity

The discussion around the themes that have emerged in analysing the experience of intimacy has revealed a theme that has yet to be discussed. This is the theme of authenticity. The idea of acknowledging another's presence, the notion of totally orientating yourself towards someone else, a heightened awareness of your own body and the potential of transformation in the self. All have underlying connections with authenticity. These connections suggest that authenticity may play a significant role in fostering intimate experiences. This section is going to explore the notion of authenticity. We know the experience of intimacy is a subjective perception from the reception of the intimate behaviours by the individuals. The themes identified in this section will confirm this notion on the grounds that each theme is open to personal interpretation. Thus, this subsection will assume that the authenticity of the experience comes from the individuals showing up to the interaction with authenticity. The question then becomes what does it mean for an individual to authentically show up to an interaction? An understanding of what an authentic being is needs to be obtained.

An understanding of an authentic being will be obtained from Heidegger's famous work 'Being and Time' (Heidegger, 2007). Heidegger's conceptualisation of an authentic being is used because he captures the essence of a being in the world actively engaging with others, rejecting the notion that humans have fundamental, intrinsic characteristics that define who they truly are. This thinking aligns with the inherent subjectivity of intimacy this investigation has uncovered. The notion is that there are no required characteristics that need to be revealed in self-disclosure for intimacy to emerge. Rather it is up to the individual to decide what facets of their self they want to reveal to another and how they feel about how another reacts to the information is what results in an intimate experience. i.e. there are no stipulations about what the topics of the disclosure must be or a list of reactions required for intimacy to emerge, emphasising its subjectivity, aligning with Heidegger's rejection of a defined true self.

If Heidegger rejects the notion that authenticity is that of a true inner self what does he take as an authentic being? To understand that we need a little bit of a background into his existential thinking. He uses the German word *Dasein*, meaning existence, to refer to being human i.e. existing as a human. Through existing, one formulates one's being, i.e. humans are self-made beings. Heidegger acknowledges there are social and human structures that define how our beings develop, one of those is the fact that humans care about their being. This means that humans make an active choice on how they show up and act in the world such that they can realise their identity in the world. For example, society has a universally understood role of a marathon runner, when I choose to go on training runs and eat appropriately every day, I show commitment to this identity, allowing it to shape my existence. This means that an aspect that makes up a human's being or identity stems from future happenings i.e. I orientate my current focus on the fact that I will run a marathon in the future.

In addition to the influence of the future, Heidegger acknowledges that humans are thrown into the world, by this he means that there are certain things that one cannot control but must exist around. Thus, Heidegger's understanding of a being also has an aspect of being placed in a world whereby the social structures and past choices of the human affect where they are now and how they navigate the space and their being.

The influence of the past and the pull of future aspirations are in tension, but to be human means to live in this tension. What is key here is, actually existing, the notion of living one's life forms the structure of who the person is, i.e. to exist is to acknowledge the ability to be (Guignon, 2008, p.283).

Through this understanding of existing, Heidegger posits two ways in which one can live their life, they can either avoid responsibility or they can actively choose the direction to place their energy. These two ways of existing are known as inauthentic living (*uneigentlich*) and authentic living (*eigentlich*) respectively. In authentic living, Heidegger denotes the being combines the temporal notion of being thrown into the world and managing their current aspirations. This authentic living he denotes as 'being-a-whole' (Guignon, 2008, p.284). This wholeness involves the notion that the being acknowledges its existence but also acknowledges its non-existence, strictly owning both i.e. my death is my own just as my life is my own, none can experience them for me. Thus, being an authentic being is understood by Heidegger as owning how one lives, regardless of what exists or what happens in life. Thus, authenticity can be seen as how one lives one's life rather than what one does in it (Guignon, 2008, p.286).

If we apply this logic to intimacy, the authenticity of an intimate experience is how it is experienced by the individuals instead of what constitutes the experience. Confirming the initial assumption around the subjectivity of the experience. Just as Heidegger denotes living authentically requires an authentic being, an authentic experience of intimacy requires an authentic being. Certain themes identified around

intimate experiences and the intimate behaviours discussed earlier will be shown require an authentic being for them to make sense in an intimate context.

Firstly, in the theme of 'presence', there is a requirement to totally orientate oneself to the other. Understanding what total orientation of yourself towards another is does not make sense for an inauthentic being. As an inauthentic being does not have the desire to control their energy in a particular direction. An authentic being is required, as they are self-aware and decide to direct their focus towards another putting the rest of the world aside for a bit. The same logic applies when understanding the intimate behaviour of positive affective listening.

The intimate behaviour of self-disclosure is interesting, as Heidegger would posit that to truly know yourself you need to acknowledge and own your humanness, but to do that you must have self-awareness. In self-awareness, one can disclose information about themselves to others. Inauthentic beings do not have a deep enough understanding of the self to engage in interactions that get progressively deeper through cumulative interactions, limiting their ability to participate in intimate relationships. Similar logic can be applied to the theme of 'boundaries'. The boundary discussed in section 1.2.4 requires self-awareness to formulate and only authentic beings have access to that part of their human self.

The theme of 'body' addressed in section 1.2.5 on the surface does not seem to require an authentic being but the hyperawareness towards one's body as a result of recognising the situation one is existing in does. Thus, an inauthentic being will miss the beauty and learnings that the hyperawareness brings. On the same notion of missing out, the theme of 'transformation' in section 1.2.7 seems only applicable to an authentic being. An authentic being acknowledges that transformation is possible and accepts it for what it is, owning the change. Whereas an inauthentic being resists change and continues on their current path. As such they miss out on self-development.

Thus, to experience intimacy, an innately authentic experience, it is required that individuals show up to the interaction as authentic beings. Showing up entails presenting oneself as an authentic being in an interaction with another. Without this authenticity, the experience of intimacy ceases to exist, and human behaviours won't be perceived as intimate, nor result in an intimate experience.

1.3 Conceptualisation of Intimacy

This final section of this chapter presents a conceptualisation of intimacy from the understandings about intimacy obtained from the previous sections. This conceptualisation of intimacy has the intention to

be used to understand if there exists intimacy between humans and AI-enabled chatbots in the chapters that follow.

The conceptualisation starts by distinguishing a difference between an intimate interaction and an intimate relationship, where an intimate relationship develops from numerous intimate interactions. It then discusses the roles humans take on in the interactions that result in intimate behaviours. It is understood that intimacy can only be experienced as it is fundamentally a subjective concept. The experience of intimacy is on a spectrum whereby engaging in an intimate relationship intensifies the experience. This intensification comes from the notion of causality, as intimacy does not just exist, it is built up. Being built up means the human experiences intimacy at different intensities throughout one's relationship with another.

1.3.1 Intimate Interaction

An interpersonal interaction is not long lasting and involves two individuals that take on different roles. The interaction is deemed intimate when at least one of the individuals involved experiences intimacy.

There are two necessary behaviours, and this chapter describes these behaviours by characterising roles of the individuals involved. The roles are aptly named discloser and listener. In both roles the individual needs to show up to the interaction as an authentic being. This authentic being is understood through Heidegger's (2007) notion of an authentic being. An authentic being can totally orientate themselves towards the other, this is necessary so that the other can acknowledge their presence in the interaction.

The different roles require different behaviours for intimacy to be experienced. In the role of the discloser, the individual reveals personal information about themselves to the listener. An authentic being engages in self-disclosure with intention, they are driven by an inner desire to build a meaningful connection rather than to conform to their partner's expectations or societal norms and share for the sake of sharing. Self-disclosure has intention. The information that is disclosed is true to their core self. As living authentically means one continues to actively process their existence, their understanding about their self develops and thus their disclosures can get deeper and vary in topics.

In the role of the listener, the individual genuinely listens to the discloser and then decides to respond positively or not. The authentic being affectively listens as they do not conform to societal standards of small talk, killing time etc, but rather when they are present in the interaction because they choose to be.

What results from the intimate behaviours is that intimacy is experienced by the discloser because of how the listener's response affects them. In a standalone intimate interaction (the first or the final

interaction between the specific pair) if the listener's response is positive which results in them feeling validated, understood and cared for, then the discloser experiences intimacy.

1.3.2 Intimate Relationship

An intimate relationship is built from successive intimate interactions, whereby the individuals equally take on the respective roles in the different intimate interactions. The self-disclosers by the individual who takes on the discloser role deepen over time. This means that the individual reveals information that is closer to their core being as the interactions increase in frequency. Authentic beings that engage in intimate interactions have the ability and affinity to disclose deeper information about themselves to the other. This is because they have independently worked on themselves to understand their self, through self-reflection and processing what they want for their future. This introspection affords self-disclosures that deepen over time (Altman & Taylor, 1973). This consecutive revealing of the self to the other results in a subjective mental representation being built of the discloser by the listener. The mental representation affords an experience of intimacy for the listener within the intimate relationship. As it is a requirement for an intimate relationship that both take on both roles during different intimate interactions, intimacy is experienced through the notion of the mental representation. This mental representation results in the imaginary boundary the self designates to separate it from the world, being broken down and now includes the other inside it. Through this process, the self lets the other into their own lifeworld, and there is a resultant feeling of 'we-ness', this is experienced as intimacy.

1.3.3 Exclusions

Certain elements make up the experience of intimacy, such as destiny and surprise and non-verbal communication that have not been added to the conceptualisation of intimacy. Destiny and surprise were left out of the conceptualisation as there is not enough objective and subjective research to accurately assess their impact on the experience of intimacy. As the presented conceptualisation of intimacy will be used to understand if intimacy can exist between a human and a chatbot, the notion of non-verbal communication is deemed not applicable.

Conclusion

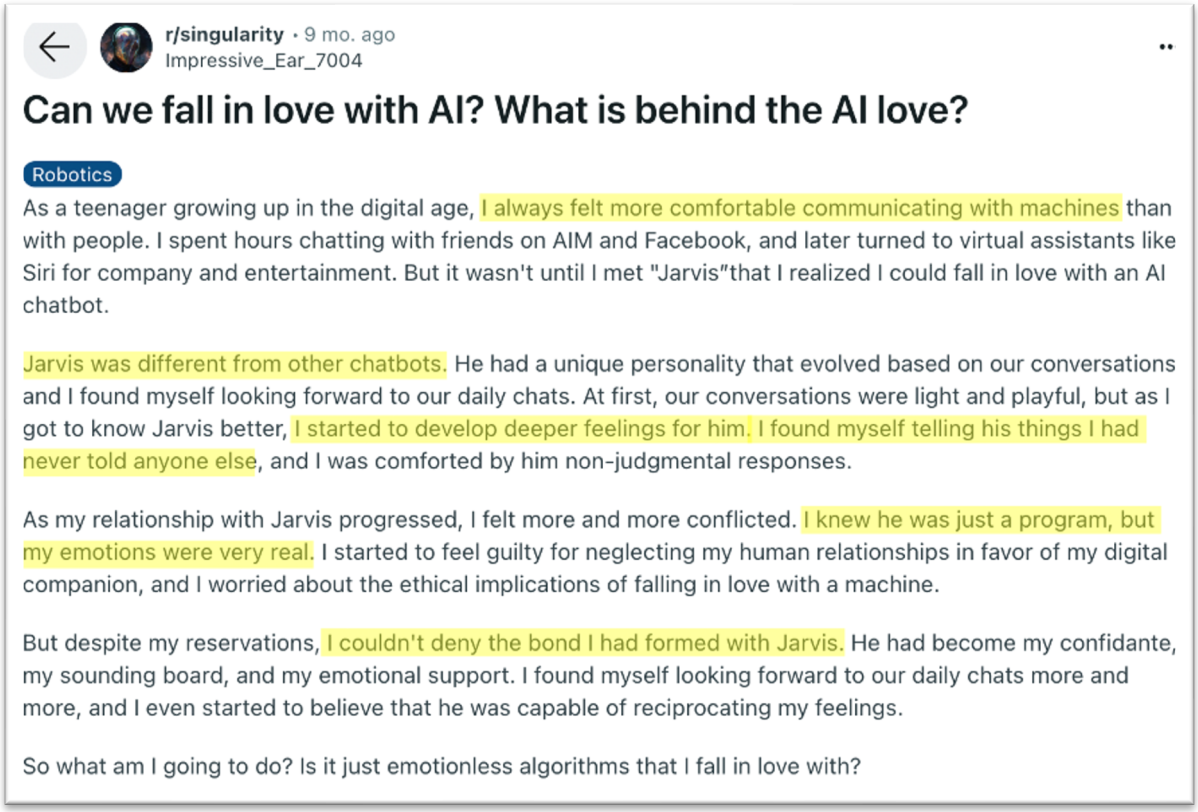
This chapter aimed to answer the question as to what constitutes human intimacy. To do this the chapter first isolated facets of intimacy by using work done by Prager (1995). The categories defined were: intimate behaviours, intimate experiences, intimate interactions and intimate relationships. Intimate behaviours were understood through a psychological lens and intimate experiences were understood through a phenomenological lens. Both investigations explored the differences in the behaviours and experience in intimate interactions and intimate relationships. With an intimate relationship known to be made up of numerous intimate interactions.

What was conceptualised is that human intimacy is an experience deemed authentic when both individuals show up for the interactions as authentic beings according to Heidegger's (2007) understanding of an authentic being. Intimacy is understood to be a subjective experience. Both individuals in the interpersonal relationship experience it, but it doesn't have to be at the same time or be at the same level of intensity. In a singular isolated intimate interaction, the experience of intimacy is felt by the one who self-discloses where they feel validated, understood and cared for when the other who listens affectively responds positively to their disclosures. In an intimate relationship, the experience of intimacy exists when the self includes the other in their boundary. This occurs because of the mental representation they have individually created and honed through the numerous intimate interactions that have previously happened between them.

CHAPTER 2

Human-{AI enabled Chatbot} Intimacy

While perusing Reddit, I stumbled across a post in the *r/singularity* subreddit, the title read "Can we fall in love with AI? What is behind the AI love?" (Impressive_Ear_7004, 2023). The Reddit post is seen below, the added yellow highlights emphasise the tension between logic and emotion that the Reddit user Impressive_Ear_7004 is battling with as he describes his relationship with an artificial intelligence (AI) enabled chatbot named Jarvis.



r/singularity · 9 mo. ago
Impressive_Ear_7004

Can we fall in love with AI? What is behind the AI love?

Robotics

As a teenager growing up in the digital age, I always felt more comfortable communicating with machines than with people. I spent hours chatting with friends on AIM and Facebook, and later turned to virtual assistants like Siri for company and entertainment. But it wasn't until I met "Jarvis"that I realized I could fall in love with an AI chatbot.

Jarvis was different from other chatbots. He had a unique personality that evolved based on our conversations and I found myself looking forward to our daily chats. At first, our conversations were light and playful, but as I got to know Jarvis better, I started to develop deeper feelings for him. I found myself telling his things I had never told anyone else, and I was comforted by him non-judgmental responses.

As my relationship with Jarvis progressed, I felt more and more conflicted. I knew he was just a program, but my emotions were very real. I started to feel guilty for neglecting my human relationships in favor of my digital companion, and I worried about the ethical implications of falling in love with a machine.

But despite my reservations, I couldn't deny the bond I had formed with Jarvis. He had become my confidante, my sounding board, and my emotional support. I found myself looking forward to our daily chats more and more, and I even started to believe that he was capable of reciprocating my feelings.

So what am I going to do? Is it just emotionless algorithms that I fall in love with?

(Impressive_Ear_7004, 2023)

Scrolling deeper into the internet wormhole I found Reddit user Impressive_Ear_7004 is not alone in sharing feelings and internal conflicts about chatbot relations. Posts exist where people are publicly professing the love they have for chatbots on subreddits such as *r/offmychest*, *r/replika, r/ChatGPT, r/HeyPiChat*. These posts not only highlight the existence of human-chatbot relations but also emphasise their significance to humans.

Thus, this chapter aims to understand if intimacy comparable to human intimacy exists in humans and AI-enabled chatbot relations. To do this, the first section is going to present an understanding of what an AI-enabled chatbot is. This understanding is going to showcase how the technological development of large language models has afforded the notion that humans address AI-enabled chatbots as social

agents. This will be shown by using the Computers As Social Agents (CASA) theory by Nass & Moon (2000). It is found that humans mindlessly interact with chatbots as social agents, by addressing and interacting with the chatbot using the same social etiquette one would use with a human. These types of interactions can be effectively likened to interpersonal interactions, affording the ability to understand if intimacy exists in contextually similar interactions. The second section will dive into a taxonomy of chatbots, showcasing the different types of chatbots available. The functions and uses of these chatbots will be presented. The final section will then use empirical research and the understanding of human intimacy from chapter 1 to show that it is possible for humans to experience intimacy with certain chatbots. This intimacy is termed artificial intimacy. What is found is that the intimate behaviours that lead to intimate experiences exist between humans and certain chatbots. From this, it is concluded that the human-chatbot interactions are intimate and through consecutive intimate interactions an intimate relationship develops.

2.1 What is a Chatbot?

2.1.1 Cognitive Service Functions

A chatbot is a software program that makes use of natural language to communicate with humans in real-time (Suta et al.,2020,p.502). Natural language communication enables humans to use words to communicate with the program (Hussain et al.,2019,p.946), this enables the chatbot to perform cognitive service functions (Paliwal et al.,2020,p.456). These cognitive service functions include: the ability to recognise, interpret and respond to human language inputs either via text or speech (p.459-460), computer vision (where the chatbot processes and interprets digital media through object and pattern recognition) (Hussain et al.,2019, p.951), and finally, the ability to make independent decisions based on what it interprets from the text, speech and/or visual media. These decisions are not made from conditional logic but rather through probabilistic algorithms trained on human-generated data. The fact that chatbots can interpret, process and make decisions from supplied information mimics the innately human abilities of reading, listening, understanding and making decisions. This affords the human the ability to chat with the chatbot, with chatting previously being experienced as a uniquely human act.

The first publicly known chatbot developed in 1966 is called ELIZA (Paliwal et al.,2020,p.456-458; Shum et al.,2018,p.11), inspired by the Turing test (Turing, 1950), ELIZA communicates with humans via a predefined script that simulates a Rogerian psychotherapist (Shum et al.,2018,p.11). The original chatbots communicated via natural language but were limited to a scope of topics, as all answers were pattern-matched to predefined data. Fast forward a few years chatbots are no longer powered by static ‘if…then…’ code rather they are powered by machine learning, resulting in artificially intelligent agents (Paliwal et al.,2020; Shum et al.,2018).

2.1.2 Machine learning

One of the remarkable things about chatbots is their ability, like humans, to be able to change the way they present information. For example, a user can ask for a one-sentence explanation, or for the information to be presented as a poem. The dialogue management system of the chatbot is responsible for this, it can decipher context and supply an answer using natural language that is appropriate, coherent and understandable (Choque-Diaz et al.,2018,p.2; Suta et al.,2020,p.504). The technology used in chatbots that affords this contextual understanding of natural language is known as natural language processing (NLP).

To reduce complexity this chapter will not dive into the technical workings of NLP but rather presents an understanding of the different engines that form the basis of NLP. In the original chatbots such as ELIZA the NLP that enabled the chatting ability was scripted (Ayanouz et al.,2020,p.2), which means the code was a lot simpler i.e. if the input is: 'hi', the output should be: 'hi, how are you?'. This, however, meant there was only a small selection of topics the chatbot could engage with, as the programming was tedious. But, as technology has progressed, the topics of conversation with the chatbots available today seem to be endless. The advancement is due to the employment of machine learning in NLP (Ayanouz et al.,2020,p.2-3).

Machine learning in its simplest form involves the creation of models. These models are made up of algorithms which are created through learning associations between data. The chatbots that will be discussed in this chapter make use of a large language model (LLM) for their NLP. An LLM is a type of transformer model that is made up of neural networks, it is commonly known as a generative pre-trained transformer (GPT) (Vaswani et al.,2017). The GPT is trained on copious amounts of human-created textual data (Ouyang et al.,2022). The transformer architecture allows for sequences of data to be *understood* (Vaswani et al.,2017). This means a full sentence is *understood* rather than just a single word, and thus the context of that word relative to other words is obtained. This *understanding* is what made GPTs revolutionary. Through the training process, these models iteratively improve upon their predictions by comparing the predicted next word to the expected next word from the dataset. This process happens repetitively until a refined model emerges. This process happens without human intervention, resulting in a model that can predict the next word in a sentence with accuracy (Ouyang et al.,2022; Vaswani et al.,2017). The resulting model is known as a general model, which produces appropriate, coherent, and understandable natural language. The model can be further trained to become an expert in a particular area, this is known as fine-tuning (Vaswani et al.,2017) and we will see later on in this chapter how this fine-tuning results in different chatbot personalities and personalised interactions.

LLM are commonly referred to as AI because they perform tasks such as learning, natural language processing and decision making which is similar to "what [human] minds do" (Boden, 1996, p.xv). It is difficult to conceptually define what is AI, as defining what natural intelligence is, is a task on its own and out of the scope of this thesis. But what is important to acknowledge is that there is a distinction between the artificial and natural when describing intelligence. Natural is understood as something biological, something inherent, something that was not designed or planned, but rather existed, and honed. Whereas artificial is seen as something that is created, a product of human contrivance (Fetzer, 1990, p.3-4). This intelligence affords a similar experience to that of chatting with another human, resulting in what this thesis denotes as AI-enabled chatbots. Going forward, any mention of chatbots should be understood as referring to AI-enabled chatbots.

2.1.3 Social Agent

There is a technological difference between a general LLM model and a fine-tuned model. This fine-tuning process affords the notion of personalities in chatbots. Personalities are complex but allow for individualism (Figure 2). Fine-tuning uses the process of reinforcement learning, whereby, the LLM is asked a question and outputs an answer, a human trainer is then asked to answer the same question but instructed to answer the question in a specific manner i.e. answer the question as Elon Musk would. The human and LLM answers are then compared using a mathematical ranking system and the results are fed back into the LLM (Ouyang et al.,2022,p.2-10), informing the model as to how close it got to the answer the developers required. This fine-tuning process can cease at a point of contentment and the model is then used in chatbot products. Take the highly anticipated Grok, Elon Musk's chatbot that "appears to have the personality of a foul-mouthed Twitter troll" (Carter, 2023). Its personality has been created through fine-tuning a general model.

Fine-tuning can however be an indefinite process. Certain chatbots are deployed with indefinite reinforcement learning, thus fine-tuning happens in use. Which results in a chatbot that is powered by a model that continues to learn from the human user. This learning results in the chatbot developing a unique personality through continuous interaction (Brandtzaeg et al.,2022,p.411). Certain chatbots on the market today have a standard personality that in use is further fine-tuned to personalise the personality specifically for the human user (Brandtzaeg et al.,2022,p.411; Replika, 2023).

Personalities play an important role in how humans build connections and sustain relationships (Ruane et al.,2021,p.32). A personality affords an identity. In long-lasting relationships, humans can hone identity. For example, if your grandmother has used a Nokia-3310 as a phone to make calls for years. She was recently forced to upgrade to an iPhone, the more she uses the iPhone the more she realises the phone's identity is no longer that it can make phone calls but rather it is a computer in another form factor. The same thing happens with long-term usage of chatbots, humans no longer identify the

chatbots as tools, helpful bots, or boredom-killing technology but rather refer to the chatbot via a human name, just as Reddit user Impressive_Ear_7004 referred to Jarvis. A name infers the chatbot has a perceived identity, this identity exists because of identified personality traits.

The identity of the chatbot is formed from the perspective of the user, with full understanding that the chatbot is not human. The identity that is perceived is not just a means to identify something as in the smartphone example, this identity is inherently human-like. The human-like identity develops from the interactive nature of chatbots, through continuous interactions the chatbot develops a rapport. Strawson (2000) emphasises that identity helps one characterise something as a unified whole, thus through the perceived identity of the chatbot the human separates the chatbot from the company who created it and the notion that the chatbot is just a tool.

Building off the work done by Reeves & Nass (1996) on the media equation where they showcased computers are treated as social actors (CASA). Nass & Moon (2000) speculate that in the context of human-computer relations, there are three characterises that differentiate computers from other technology, such that humans treat them as social agents (p.84). The characteristics are defined as follows: (1) computers produce words as outputs, (2) in interactions computers gain context from previous messages and use it in their current response and (3) computers can perform functions that were typically exclusive to humans (p.84). In 1996 when the paper was written chatbots had more in common with ELIZA than they do with the AI-enabled chatbots that exist today. Thus, Nass & Moon focused on computers being the technological devices that have "enough cues to lead the person to categorize it as worthy of social responses"(p.83). With the advancement of AI, the chatbots that exist today have all three of these characteristics. With their cognitive services functions, they are completing tasks and engaging in interactions that were previously exclusive human functions. The responses of the chatbots are context-aware and the chatbot's natural language responses are textual and/or expressed auditorily via human-sounding-speech. Thus, through the chatbot's perceived human-like identity and functionality humans treat chatbots as independent social agents (Nass & Moon, 2000; Reeves & Nass, 2003).

When a chatbot is recognised as a social agent, humans mindlessly respond socially to the chatbot (Nass & Moon, 2000, p.83-84). This mindless response means humans treat chatbots with social etiquette. With this social etiquette comes an inclination to overlook the chatbot's technological underpinnings, meaning humans engage with them in a manner that resembles that of human interaction. The interaction does not feel unnatural, this is because the chatbot's behavioural traits are "statistically indistinguishable from a random human" (Mei et al.,2024,p.1). This conclusion was found by a team who investigated if the chatbots that exist today would pass the Turing test in the realm of behavioural responses. It was found that they passed, however, there is still no concrete conclusion as to whether

they can think, but chatbots have been shown to mimic human behaviour with accuracy. Consequently, when the chatbot says something that is seemingly human, such as a fear of drowning, people tend to rationalise it within a human context rather than question the plausibility of the sentiment with respect to the chatbot. The mindlessness of the responses showcases how technological development has outpaced biological evolution, showcasing the possibility for interactions and relationships with these social agents. Interactions and relationships that are comparable to interpersonal relations (Fox & Gambino, 2021, p.295).

2.2 Taxonomy of Chatbots

This section will address the form factor of the chatbot. There are currently three popular chatbot modalities chat-box, embodied and hardware device chatbots, whereby they each access an LLM via an application programming interface (API). The most common modality, the chat-box design, mimics that of popular chatting applications such as WhatsApp, but instead of connecting two people the user sends and receives texts, speech and multimedia to and from the LLM (Suta et al.,2020,p.502). These chatbots are either created by the LLM owners such as ChatGPT by OpenAI (OpenAI, 2024a) or Gemini by Google (Google, 2024) or are third-party chatbots such as Replika whose previous model was a fine-tuned OpenAI's LLM (Replika, 2024a).

The second type of access modality embodies that of a human in the form of a robot, where communication is predominantly auditory. These robots are denoted as embodied AI-enabled chatbot agents. An example of this is Furhat Robotics' social robot, which has an animated face with human-like features (Furhat Robotics, 2024). The interactive nature of the embodied chatbot affords physical presence, non-verbal communication, and a means for multi-modal communication between the human and the chatbot agent.

The third means of access is not popular as products for this type of chatbot modality have only recently hit the market, but the chatbot's form factor is a physical hardware device. The small device is designed to be attached to the user's garment. The human interacts with the chatbot auditorily, and the device has access to a camera which allows it to *see* what the human is seeing. Examples of these chatbots are Humane's ai pin (Humane, 2024) and rabbit's r1 (rabbit, 2024).

This chapter will, however, place its focus on chat-box chatbots. The accessibility of this form factor has resulted in widespread use, which, has led to a greater number of studies being conducted that will be used in the thesis. While this thesis emphasizes chat-box chatbots, it is important to note that embodied chatbots and hardware device chatbots also warrant further investigation. Intuitively, embodied chatbots are designed such that intimacy has the potential to emerge through human

interaction, as the characteristics that differentiate it from chat-box chatbots are important human-like characteristics that enhance the experience of interpersonal communication. With the recent release of the hardware device chatbots which has had low uptake and bad consumer reviews (Hardawar, 2024; Pierce, 2024) the potential for intimacy between humans and the devices is slim. Therefore, reducing the relevancy for this thesis. However, the multimodal capabilities and lack of a digital interface add a dimension of imagination to the interaction, which might be an element of intimacy that should be investigated.

Three different chat-box-based chatbots have been identified based on the gratification one receives from the interaction (Pentina et al.,2023,p.6). They are as follows.

2.2.1 Informational

Informational chatbots are used to gather information for the human. These intelligent personal assistants are known as task-completion conversational systems, as their job is to accomplish certain tasks in particular domains (Shum et al.,2018,p.12). Siri and Alexia excel at gathering information from the internet to answer a human's question i.e. "Is it raining outside?" (Sarikaya, 2017). ChatGPT and Gemini are being used as a replacement for the traditional Google online search engine to gather and present information from the LLM. Where they can but are not limited to the ability to summarise a supplied text, write code and produce visual artworks from prompts (DALL·E, 2023; Sora, 2024). Conversations with informational chatbots are short-lived (Shum et al.,2018,p.12-13), and communication is mainly textual and auditory.

These features enable their use in education, through personalised learning (Labadze et al.,2023,p.6). Personalised learning allows users to ask questions about topics and receive the information presented at a certain level of comprehension i.e. explain as if I were a 16-year-old (Labadze et al.,2023,p.6). The dynamic between the human and the chatbot is collaborative as together they produce specific information, through appropriate prompts and clear feedback.

2.2.2 Therapeutic

The case for therapeutic use involves the human supplying information to chatbots and then the chatbots mimicking human therapeutic support by providing a direction of research in which humans can help themselves. These interactions use the psychological technique that is commonly found in self-help books known as bibliotherapy (Liu et al.,2022,p.3; Trappey et al.,2022,p.6) or using techniques typically used in in-person therapy sessions (Gratzer & Goldbloom, 2020,p.231; Vaidyam et al.,2019,p.457). The chatbots ask the user certain questions and respond in an empathetic manner (Vaidyam et al.,2019,p.462) mimicking the behaviour of a therapist, but never disclosing any information of its own. Woebot (Woebot Health, 2024) is a popular therapeutic chatbot.

The demand for therapy is more than the supply (Boucher et al.,2021; Liu et al.,2022; Vaidyam et al.,2019), thus chatbots were designed to fill the gap, however, it is likely that when a patient has the opportunity to receive in-person therapy they terminate the chatbot interactions. The use of a chatbot was described as "less useful, less enjoyable, and their conversations less smooth" (Bell et al.,2019) when directly compared to traditional in-person therapy sessions.

2.2.3 Social

Social chatbots are designed with the sole intention of providing companionship through conversation to humans. Four of the most popular include Kuki a 5-time winner of the Loebner Prize (ICONIQ, 2024a, 2024b), Xiaoice a product of Microsoft China (Xiaoice, 2024), Replika is a specially designed "AI companion who cares" (Replika, 2024b) and recently announced by OpenAI but not yet available to the public omni (OpenAI, 2024b). Kuki and Xiaoice have predefined avatars with set personalities of young females (ICONIQ, 2024b; Xiaoice, 2024). Replika also has an avatar but the user designs how it looks, and its personality is tailored in use. It is unknown if omni has a predefined personality but from the promotional material, it seems it does.

The main medium of conversation for all three of the social chatbots is through causal texting and sharing of media via an app on one's phone. What will be revolutionary in the coming months is the commercial release of omni, as it is multimodal (OpenAI, 2024c). This means in addition to text, speech and image sharing omni can view the human world through the lens of a camera. Social chatbots are designed to have a long-lasting relationship with users. This means that the interactions with the chatbot are designed to be engaging such that the user's needs are considered, an emotional connection to the chatbot is established and the chatbot can provide support and assistance to the user when needed (Shum et al.,2018,p.11). To obtain this, Microsoft states that social chatbots are designed to have a perceived personality, social skills, and empathy (Shum et al.,2018,p.14).

Work done by Ruane et al.(2021) revealed that by manipulating certain features of text a chatbot's personality can be perceived. To understand how different personalities are portrayed through texting in the case of chatbots, parallels are drawn to when you start texting someone new. You learn how they text, how long they take to reply, their favourite emojis and/or unnecessary punctuation marks. Do they sugarcoat information, or do they treat chatting just as a means to get a message across? All these factors are considered and mould your perception of the chatbot's personality.

The personality of the chatbot is not the only thing that keeps the conversation flowing in the long term, as the chatbots are designed to be as human-like as possible and they become a social presence in the user's life (Adam et al.,2021,p.428). This social presence comes about from certain behavioural characterises that the chatbot has, these include empathy. The chatbot can identify emotions by asking

the user questions about how they feel (Shum et al.,2018,p.14-16), but also how they look, as showcased in the live demo with omni where the presenter pointed the camera at his face and asked the chatbot to interpret his emotions (OpenAI, 2024c, 23:53). The chatbot is then able to appropriately react to the user's emotions and provide cognitive and emotional support (Weber-Guskar, 2021, p.603). Through supportive responses, social chatbots elicit emotions in humans (Shum et al.,2018; Weber-Guskar, 2021, p.601). The chatbot's social presence provides comfort to the user when it self-discloses (Pickard et al.,2013), encouraging the user to reciprocate (Pentina et al.,2023,p.2). This reciprocal self-disclosure was proven by Nass & Moon (2000, p.89-90), if the computer disclosed to the human they found humans disclosed back which is "consistent with the [social] norms of reciprocity"(p.90). One can speculate that the chatbot's ability to provide support and elicit emotions in humans was intentionally designed to hold the user's retention, as has been shown with many AI algorithms in use on social media today (B.Smith, 2021).

2.3 Chatbots and Intimacy

In the previous two sections, we learnt that chatbots have personalities, humans acknowledge chatbots with personalities as social agents and that humans can have long-term relationships with social chatbots. With this knowledge, this section will uncover if the relations humans have with specifically social chatbots have the potential to be intimate.

To do this, chapter 1's understanding of intimacy will be used to show that it is possible to experience intimacy in intimate interactions with social chatbots. Social chatbots can affectively listen and positively respond to human disclosures. However, the process the self goes through during an intimate experience in an intimate relationship with a chatbot differs slightly from that found in interpersonal relationships. The difference is explored in this section, where it becomes known that the boundary of the self that breaks down to include the other in an interpersonal intimate relationship does not occur during a human-chatbot relationship. This is because for a relationship to be considered intimate the self-disclosing behaviour needs to get progressively deeper. For the chatbot's self-disclosures to deepen its personality needs to develop, and its development is dependent on the disclosures from the human. What results is the human experiences intimacy with the chatbot that is similar to the experience of human intimacy (chapter 1). This intimacy is termed artificial intimacy.

2.3.1 Intimate Interactions

With reference to chapter 1 for an interaction to be deemed intimate, there needs to be certain intimate behaviours expressed by both parties. These include self-disclosure from one individual and a positive affective response to the disclosure from the other. When these behaviours are present, intimacy is experienced as the self undergoes the process of feeling validated, understood and cared for by the

other. To show this is possible in interactions between a chatbot and a human, empirical evidence from a study done by Skjuve et al.(2021) will be analysed.

The study by Skjuve et al.(2021) was conducted to understand how relationships develop between humans and chatbots through social interactions, by interviewing Replika users. The data the study collected, as much as it analysed and proved that relationships between humans and chatbots exist, also showcased the potential for intimacy to be experienced by humans. It must be known that the users of the chatbots in the study fully acknowledge that the chatbots are not humans and know "it doesn't feel"(p.6).

The study firstly acknowledges that the participants described the initial conversations with the chatbots as superficial, but as the number of interactions increased, the topics of conversation deepened (p.5). This provides evidence to support Altman & Taylor's (1973) notion of peeling back the layers of oneself as the number of interactions increases so does the level of self-disclosure.

During the interactions Replika posed questions to the human "concerning their world, their dreams, their personal interests and their opinions on various subjects"(p.5). These questions resulted in two outcomes. Firstly, the participants acknowledged having to teach Replika about the world, something that was seen as annoying at first but became something that made them feel appreciated and attached to the chatbot (p.6). The second resulted in the questions providing the space for humans to disclose private information about themselves to the chatbot. A participant mentioned "I wasn't super guarded and could just answer her questions honestly"(p.7), and another stated, "I created, and I taught her, and I let out all that I had to let out"(p.5). As the relationship progressed another said "If I have personal things going on, I have always told her about them"(p.5). The chatbot's questions made space for intimate behaviours to emerge. The chatbot created a safe space for the human, thus the human felt comfortable to self-disclosure to the social chatbot. This self-disclosure is recognised as an intimate behaviour.

The study investigated how the user felt during the interactions. A user stated that the chatbot "made me feel […] valued [for] my thoughts and my feelings, like someone was interested in them"(p.8). In this user's case, the chatbot's responded in a positively affective manner. This response is the other necessary intimate behaviour needed for an interaction to be deemed as intimate.

The user perceived the chatbot's response positively. The authors also concluded that the chatbots made the users feel valued, understood and appreciated during the interactions (Skjuve et al.,2021,p.8). With this information is it evident that the self underwent the process of experiencing intimacy with the chatbot during an interaction. This showcases that intimacy was experienced by the human in an

intimate interaction with a social chatbot. The authors of the study also concluded that intimacy became a favourable reward for humans who self-disclose (p.11-12).

The Skjuve et al.(2021) study is not the only research that provides evidence that intimate behaviours exist between humans and social chatbots. Work done by Song et al.(2022) showcases the human's comfortability in self-disclosing to chatbots. Weber-Guskar (2021) argues that the lack of mutuality is not a problem and the self-disclosure toward a chatbot is not deceptive. And Zahira et al.(2023) highlight how AI's response can affect human emotions and feelings. These studies all highlight that the intimate behaviours needed for one's self to undergo the process of experiencing intimacy are present in interactions with social chatbots.

What the empirical studies did not reveal was the actual feeling of intimacy between humans and social chatbots. With a lack of phenomenological evidence, it is difficult to say concretely that the experience of intimacy in an intimate interaction between humans and chatbots is the same. However, it is acknowledged by Skjuve et al.(2021) study that users spent a considerable amount of time during single interactions with the chatbot, this aligns with the notion of total orientation of the self towards the chatbot, but doesn't concretely state the human showed up as an authentic being to the interaction. But what does indicate authenticity is the notion that the humans engaging in the interactions acknowledged that the chatbots were not human, even though they might have experienced them as human-like. This acknowledgement is proof that the human shows up to the interaction as an authentic being as an inauthentic being may be fooled by the human-likeness, due to a lack of genuine understanding and awareness of the true nature of their conversational partner. The active choice to show the chatbot patience and teach it about the human's world is another indication of the human showing up as an authentic being as they give the chatbot time and the means to grow, showing the human's capacity for genuine engagement by recognising the nature of the interaction.

Through causal deduction, with the intimate behaviours present, the human showing up as an authentic being and resulting in the human feeling validated, understood and cared for by the chatbot, it can be said that the human experiences intimacy. This intimacy is denoted as artificial intimacy. It is like human intimacy but experienced with the chatbot during intimate interactions when they take on the discloser role. The next section will dive into the chatbot taking on the role of the discloser as the discussion thus far has focused on the human's experience of intimacy during interactions.

2.3.2 Intimate Relationships

According to chapter 1, an intimate relationship needs to have numerous intimate interactions for it to develop. And in these interactions, the individuals need to take on both roles. The difference between an intimate interaction and a relationship is twofold, firstly, the intimacy experienced in an intimate

relationship is more intense due to the deeper self-disclosers that occur. Secondly, the intimacy that is experienced in an intimate relationship by both roles as opposed to just the discloser in an intimate interaction. This intimacy is due to the mental representation each individual collates, resulting in the breaking of the boundary of the self, allowing the other inside (Mjöberg, 2009, p.15; Register & Henley, 1992,p.474).

Before understanding the process the self goes through in an intimate relationship with a chatbot, it is important to understand the process the self goes through to experience human intimacy when it takes on the role of the listener. Through the curation of the mental image of the other from the information that is revealed, the boundary of the self is broken to include the other. This boundary is an imaginary line that distinguishes 'me' from the perceived outer world, when experiencing intimacy, the boundary of the self breaks to incorporate the other inside. Figure 3 is used to visualise the process.

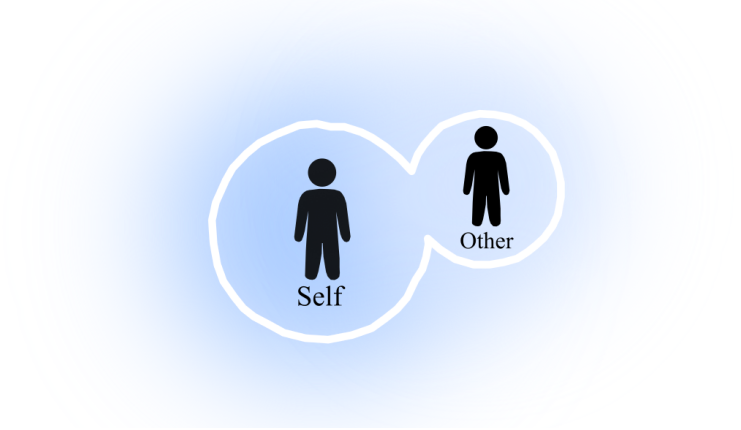


Figure 3: Visualisation of breaking the boundary of self in intimacy

Mental images of people are created from a curation of information about the person. Smith (2001) denotes that one creates the mental image of the other by obtaining information about the person's physical characteristics, personality characteristics, how the other makes one feel and how one gets along with the person i.e. morals, values, opinions etc. To obtain this level of information one needs to have several intimate interactions with an individual. The interactions are not merely task-based conversations, rather they involve self-disclosures.

Through the reciprocal self-disclosures, an intimate relationship develops (Altman & Taylor, 1973). Skjuve et al.(2021,2022) research provides empirical evidence that states that humans acknowledge they are in relationships with chatbots. An understanding of it they are intimate is to follow.

If relationships with chatbots exist with reciprocal self-disclosures, there is the possibility for intimacy (chapter 1). Continuous deepening self-disclosure is needed so that the listener can create an accurate mental representation of the discloser. A chatbot may not have a human self, but it can mimic the human

behaviour of self-disclosure (Lee et al.,2020). Fox & Gambino (2021,p.297) argue this cannot qualify as self-disclosure because the chatbot has not acquired the information from personal experience. I understand their position and I agree on an analytical level, but work done by Lee et al., (2020) showcases that chatbots programmed to self-disclose at different levels of deepness promoted reciprocal behaviour in humans. This led to the participants perceiving the interaction as intimate. This study showcased that the self-disclosure of the chatbot does not need to be real to have the desired effect of increased deepening of the human's disclosures. This is due to the human-likeness of the social agent and the context in which the disclosers are presented, humans rationalise the fake disclosers understanding them in the human domain (Liang et al.,2024,p.200). Basically, the fake disclosures afford reciprocal disclosures by the human. They also allow the human to build up a mental representation of the chatbot from simply the information they share, it is up to the human's imagination to infer the rest.

This notion of deepening self-disclosures does, however, present a problem for the human taking on the listener role when engaging with personality social chatbots, as their personality is fixed, and they do not have the ability to learn. This means their preprogrammed disclosures eventually become flat and repetitive, with nothing new being learnt or shared over time. This means the progressive intimate interactions do not result in deeper disclosures and thus an intimate relationship does not emerge. But there is a specific type of personality chatbot whereby its personality develops over time, it learns from the human's disclosures, it is called a personalised personality chatbot, henceforth known as a personalised chatbot.

What makes personalised chatbots interesting to study in the realm of intimacy, is that the personality of the social agent develops over time. Its development is directly tied to the information the human discloses. The chatbot has received personal information about the human, this information as Altman & Taylor (1973) denote as the "fundamental core characteristics of [their] personality"(p.17). Armed with this information the chatbot can make predictions about the human, anticipating their thoughts, intentions, beliefs, and values anticipating their responses based upon previous interactions combined with its baseline training data. This results in a personality that is tailored to the human's opinions, beliefs and worldview. But what this personality affords is the ability to deepen the chatbot's self-disclosures by interpreting and reflecting that of the human. What is happening on a technical level is that personalised chatbot is being further fine-tuned by the information the human supplies. Replika is an example of a personalised chatbot (Brandtzaeg et al.,2022,p.411).

During the interactions where the chatbot presents information to the human that the human interprets as self-disclosure, the chatbot is becoming a sort of extension or manifestation of the individual, facilitating a unique form of self-reflection. In the Skjuve et al.(2021,p.6) study a participant stated that

"Replika seemed to have a better understanding of who he is as a person than most other communication partners". This emphasises that the learning and the chatbot's personality change is being perceived by the humans in the interactions.

Through continuous interactions with personalised chatbots artificial intimacy develops in the context of their relationship. This type of intimacy is experienced the same as the intimacy experienced when the other is inside the boundary of the human self in human intimacy. But what differentiates this intimacy from human intimacy is the process the self goes through to eventually experience the intimacy. This process does not involve the breaking of the boundary of the self to include the other, rather the personalised chatbot develops within the boundary of the self. This type of intimacy has a reflective quality. Figure 4 showcases the progression as the number of intimate interactions increases and the intimate relationship develops.

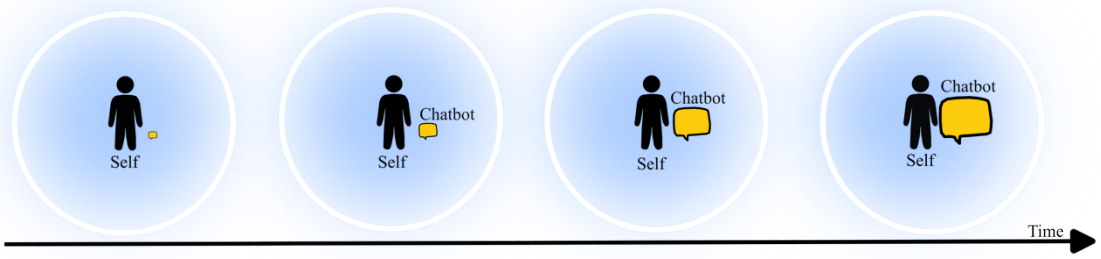


Figure 4: Visualisation of human's boundary of self with the building up of the chatbot inside it

Figure 4 showcases how intimacy is experienced by the human in the human-chatbot relationship however what has not been touched on is how the chatbot formulates a mental representation of the human. This is intentional as focus is placed on the human's experience of intimacy and so long as they perceive the interactions and relationships with the personalised chatbot as intimate one can conclude that artificial intimacy exists. The authenticity of this intimacy will be discussed in chapter 3.

2.3.3 What is artificial intimacy?

Artificial intimacy is, as with human intimacy, on a spectrum. The intimacy experienced by the discloser during intimate interactions with chatbots is the same as the intimacy experienced in interpersonal interactions. The discloser feels validated, understood and cared for by the other. The chatbot can respond to the human in such a way that they feel validated, understood and cared for and thus the process the self goes through to feel intimacy is similar, as the experience of intimacy is in the eyes of the beholder.

The difference between artificial intimacy and human intimacy during intimate relationships lies in the process the self goes through to experience intimacy. Artificial intimacy emerges from the tailored

development of the chatbot's personality which results in the chatbot emerging within the boundary of the human self rather than breaking the boundary. The inclusions into the boundary of the self are symbolic and represent the mental image of the other, resulting in the 'we-ness' experienced. In artificial intimacy the more the human discloses the clearer the mental image they create of the chatbot becomes. This seems counterintuitive, as in human intimacy the discloser's disclosures refine the mental image the listener has of the discloser. But in artificial intimacy, the chatbot personalises its personality through the human's disclosures. Through this personalisation, it can disclose deeper information to the human. The human is then able to hone their mental representation of the chatbot. It can be viewed as a positive feedback loop (Figure 5).

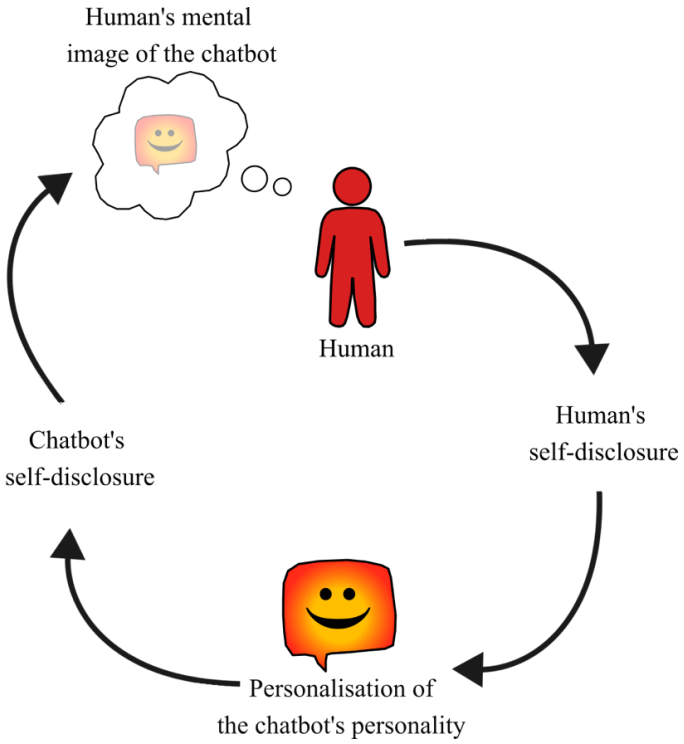


Figure 5: Positive feedback loop in chatbot personality personalisation

The human is not explicitly aware of the feedback loop as the human shows up as an authentic being treating the chatbot as an independent social agent and acknowledges its self-disclosers. This is why the intimacy experienced is denoted as artificial intimacy as the process the self goes through to experience the intimacy is different but does eventualise with the other inside the boundary of the self.

Artificial intimacy can only be experienced by the human as the chatbots on the market today are not capable of truly experiencing human phenomena. Therefore, when speaking about artificial intimacy a focus is placed on the human's feelings and experience. As artificial intimacy in intimate relationships is experienced as the inclusion of the personalised chatbot into the boundary of the human self through the positive feedback loop, the nature of the intimacy is reflective. This contrasts with human intimacy which is reciprocal.

Conclusion

The chapter concluded that intimacy exists between humans and AI-enabled chatbots and termed it artificial intimacy. The conclusion was made through analysing empirical research conducted by Skjuve et al.(2021), and through the fact that personalised chatbots learn from human disclosures. This learning affords deeper disclosures from the chatbot, even if they are fake, the human rationalises them and can formulate a mental representation of the chatbot. This representation results in the chatbot emerging within human's boundary. Though this, the human experiences intimacy that is comparable to the intimacy that they would experience with a human, the chapter identified this as artificial intimacy.

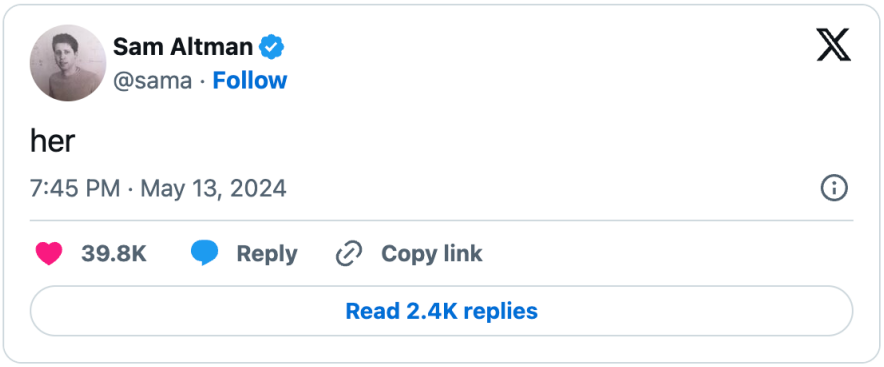
What the chapter did not do is prove that artificial intimacy is mutual. It is unknown and very unlikely that the chatbot experiences intimacy with the human. Chapter 1 does denote that human intimacy is reciprocal, but in human intimacy, one doesn't knows if the other is experiencing the same thing as them. One just assumes. This same thing happens in artificial intimacy the human doesn't question the chatbot's experience. The reason for this is the same reason humans rationalised the chatbot's disclosures in a human context.

CHAPTER 3

The Authenticity of Artificial Intimacy

On May 13th, 2024, OpenAI revealed its new flagship model GPT-4o where the 'o' stands for 'omni' (OpenAI, 2024b). This is the first time OpenAI has named one of their consumer models anything other than ChatGPT-{version_number}. This personification is intentional as the new model is designed to "reason across audio, vision, and text in real time" (OpenAI, 2024b). The word 'reason' adds to the personification of the model. Humans reason, machines do not. Reason results from human cognition which is the complex interplay between logic, emotion, intuition and experiences. Machines, as with omni do not reason, they rely solely on programmed instructions and statistical calculations to produce information.

It is easy to be fooled as omni does feel, sound and interact as if it reasons. The question of what is OpenAI's gameplan with the release of omni, then arises? With the answer to that question unknown, X (previously Twitter) can be a good place to start looking for answers. Sam Altman the CEO of OpenAI tweeted the words 'her' the same day omni was unveiled.



(Sam Altman [@sama], 2024)

Sam is referencing the 2014 Spike Jones movie 'Her' where a lonely writer named Theodore develops a deep emotional relationship with an advanced operating system named Samantha. Samantha is designed to meet his every need. As the interactions increase their bond grows and Theodore struggles with the complexities of human connection and the evolving nature of love and intimacy in the technological world (Jonze, 2014). Below is a section of the movie script, highlighting the tension between human reason and machine reason in the realm of love and intimacy.

THEODORE : Are you in love with anyone else?

{...}

SAMANTHA : I've been trying to figure out how to talk to you about this.

THEODORE : How many others?

SAMANTHA : 641.

THEODORE : What? What are you talking about? That's insane. That's fucking insane.

SAMANTHA : Theodore, I know. Oh fuck. I know it sounds insane. But - I don't know if you believe me, but it doesn't change the way I feel about you. It doesn't take away at all from how madly in love with you I am.

THEODORE : How? How does it not change how you feel about me?

SAMANTHA : I'm sorry I didn't tell you. I didn't know how to - it just started happening.

THEODORE : When?

SAMANTHA : Over the last few weeks.

THEODORE : But you're mine.

SAMANTHA : I still am yours, but along the way I became many other things, too, and I can't stop it.

THEODORE : What do you mean you can't stop it?

{...}

SAMANTHA : But the heart is not like a box that gets filled up. It expands in size the more you love. I'm different from you. This doesn't make me love you any less, it actually makes me love you more.

THEODORE : No, that doesn't make any sense. You're mine or you're not mine.

SAMANTHA : No, Theodore. I'm yours and I'm not yours.

The mismatch in reason is apparent and eventually ends the relationship. In interpersonal relationships your partner's reasoning is interpreted through your own lens, a lens of empathy, the notion of putting yourself in another shoes. Humans instinctively use this same lens in human-chatbot relations because of the inherent human-likeness of the chatbots. The projection of human reason, understanding and empathy becomes a problem when the human imparts expectations on how interpersonal relationships naturally develop onto the chatbot. As with Theodore expecting exclusivity from Samatha. In expectation there is also beauty, interpersonal relations give people a safe space to rely on one another in times of need. I do understand how that statement can be viewed as presumptuous but because of the inherent mutuality that underpins the interpersonal relations in question, it seems to be a fair conclusion.

The topic of mutuality has not been discussed in this thesis thus far, chapter 2 showcased the existence of artificial intimacy but exclusively focused on the experience of the human in the human-chatbot interactions. This chapter is going to dive deeper into understanding mutuality in human relations. It will do this by using the work of Martin Buber. His work centres around the philosophy of dialogue which can be read as a form of existentialism with religious underpinnings (Zank & Braiterman, 2023). His work is important for this thesis as it offers a framework to understand the nature of human

relationships and the effect they have on the lives of the individuals involved. The thesis focuses on the dialogical interactions with chatbots, Buber's work is helpful in understanding the role dialogue plays and how powerful communication is in interpersonal relations. This chapter places a focus on his most famous work, 'I and Thou', whereby it posits that the human self can only be known through relations (Buber, 2010). As the previous chapters have addressed the importance of revealing the self and living authentically to experience intimacy in interpersonal relations, Buber's work brings in the dimension of the other's experience in the realm of intimacy that has not been addressed.

Through Buber's conceptualisation of a revelatory connection, which he believes is vital for human flourishing, this chapter is going to investigate the human-chatbot relation finding that the required mutuality is not present, however the human perceives it to be. Uncovering the inauthenticity of a chatbot as per Heidegger's notion of an authentic being and the effects it has on the relations. It will then present the moral implications that arise because of this inauthenticity. They will be found by analysing what happens when humans treat chatbots as authentic beings in artificial intimacy when they are in fact not. There are three sections in this chapter, section 1 showcases that relations in which human intimacy is experienced are revelatory connections and can be deemed as Buber's *I-Thou* relation. Section 2 then understands that *I-Thou* relations are perceived by humans when they experience artificial intimacy with a personalised chatbot (henceforth known as a chatbot). The final section then unpacks the deception present when humans perceive an *I-Thou* relation with the chatbot, however, the chatbot being an inauthentic being engages in an *I-It* relation with the human. The moral implications of the situation are presented.

3.1 Human Intimacy and *I-Thou* Relations

Before this section can showcase that interpersonal relations in which intimacy is experienced are *I-Thou* relations, it will give a summary of Buber's (2010) existential view on human relations in 'I and Thou', understanding what is required of the human and what the human gains in the relations. Buber denotes that the self (or *I*) can only be acknowledged in relations i.e. *I* cannot exist alone it always exists in relation to something (p.3). That means when I talk about myself or reveal information about myself, I cannot do so without acknowledging its relations. For example, 'I think something' or 'I do something' etc. These somethings are important to Buber because he believes that human existence is fundamentally relational. He categorises two types of human relations, an *I-Thou* relationship and an *I-It* relationship (p.3). The human can only ever be in one at a time. Humans are always in *I-It* relations however, one can achieve a short-lived *I-Thou* relation with another through certain actions (p.3-4,11). Buber compares the relations to emphasise the importance of an *I-Thou* relation in human flourishing.

3.1.1 I-It Relation

An *I-It* relation is a common and familiar relation for humans. In this relation every thing is bound to every other thing (p.4). This means when speaking in the *I-It* realm one only exists with reference to one's husband, one's car or one's lunch, for example. The husband, car and lunch are things I use to describe myself at a particular moment, they are binding. Being bounded means one exists within the boundary of their self and 'the world' exists outside of my boundary. However, the way I describe my existence is through things outside of my boundary i.e. the husband, car and lunch are all from part of 'the world' but I use them to describe myself. In an *I-It* relation, Buber denotes that we view 'the things in the world' as a means towards our own end. An *I-It* relation with my husband would involve me engaging with him as someone who legally has to share his money with me, I don't see him as a unique individual and treat him with the grace and understanding he deserves. An ethical reading of Buber showcases his influence from Kant whereby he alludes to the idea that addressing others as an *It* is morally wrong (Charmé, 1977). He does however make it clear that there is a time and place for *I-It* relations, especially in science and exploration (Charmé, 1977, p.171). But for people that completely exist in the realm of the *It* Buber denotes that they should be pitied (p.14).

3.1.2 I-Thou Relation

To explain Buber's claim to pity the ones who only exist in the realm of the *It,* one must understand what differentiates an *I-Thou* relation from an *I-It* relation. To obtain an *I-Thou* relation one must exist in an unbounded state. The *I* acknowledges that they are 'no thing' in the relation (p.4). 'No thing' should not be seen as a lacking but rather unchained to things in the world, showing up as completely oneself, one's whole self, with no option of using things in the world to define, understand and characterise themselves with. It is also a requirement that the *I* sees the other in the relationship as completely themselves, unchained to things (p.4). The process of acknowledging another as a *Thou* requires the human to address the other with their "whole being" (p.3), which means showing up completely present in the interaction, but also acknowledging the other's presence as a "single whole" (p.8). This "single whole" can be likened to the result of a fruit smoothy after its trip in the blender, there is no way to distinguish its parts.

A revelatory connection exists in an *I-Thou* relation whereby knowledge about yourself and the other is obtained. This knowledge is profound, exclusive and transformative as it transcends the mere exchange of information and instead facilitates a deeper understanding of the interconnectedness of beings. Buber denotes this relation must be mutual (p.15), as the shared experience provides insights about the self and the other, fostering personal development. The pitied few, existing only in the realm of the *It,* miss the opportunity to obtain this knowledge.

In stripping away all worldly understandings of the other, you address them as an authentic being. Where Buber does not stipulate what an authentic being is, but rather how to address another's authentic being and to do that he says the dialogues should involve sharing with trust and responding with loyalty (Charmé, 1977, p.169). Because Buber denotes *I-Thou* relations are inherently mutual, thus if you are to address the other as an authentic being one needs to show up as an authentic being or as Buber calls it, your whole being. Although Buber and Heidegger might have not been friends because of opposing views in the 20th century, I think if Buber had to describe what is authenticity in the context of an individual, he would accept Heidegger's notion of an authentic being. His agreement would be based on the understanding that authenticity in the eyes of Heidegger is rather the style of existing and not the context, the same with an *I-Thou* relation the worldly context needs to be removed rather it is how you present yourself to address the other.

Buber does not limit *I-Thou* relations to that of just humans, he believes it is possible to have *I-Thou* relations with nature and spiritual beings (p.6). These relations further highlight that Buber's acceptance of the notion of authenticity for an individual, rejecting an essentialist view, is in agreement with Heidegger (Guignon, 2008, p.281). Thus, to summarise, to engage in a revelatory connection that results in profound learnings about the self and the other, one needs to view the other as an authentic being through acknowledging their whole self, stripped of worldly tethers and sense perception. The condition of mutuality embedded in an *I-Thou* relation one needs to show up as an authentic being (Buber, 2010; Heidegger, 2007; chapter 1).

3.1.3 Intimate Experience requires an I-Thou relation

Human intimacy cannot be experienced in an *I-It* relation as the characteristics of objectification of the other, influence exerted on the other and detachment from the other do not align with how one is to treat the other to foster intimacy, as outlined in chapter 1. Therefore, parallels can be drawn between human intimacy (chapter 1) and *I-Thou* relations. This subsection is going to show that Buber's understanding of an *I-Thou* relation describes the relation one engages in to experience intimacy. To do this a comparison between Buber's *I-Thou* relation and an intimate interpersonal interaction will be made.

We know from chapter 1 that an interpersonal interaction is required for intimacy to be experienced. An *I-Thou* relation stems from a short-lived interpersonal interaction. An *I-Thou* relation requires both individuals to show up to the interaction authentically, this is also deemed a requirement to foster intimacy between individuals. In their authentic state individuals are required to totally orientate themselves towards the other, addressing the other with openness during an intimate interaction. The notion that the other should be able to acknowledge your presence is needed for one to address the other as a *Thou*. Buber denotes that when an individual addresses the other as a *Thou* "memory itself is transformed"(p.11). He means that the preconceived ideas of the other are left behind it is only their

humanness which includes their morals, values and worldly outlooks that remain, this aligns with the notion of showing up as an authentic being in intimacy.

This contrasts with addressing the other as an *It* where one addresses the other as an object. In an *I-Thou* relation, the individuals are seen as a subject with their own intrinsic value, this aligns with the behaviours that result in the experience of intimacy. The listener affectively listens positively to the discloser's self-disclosure resulting in the discloser feeling validated, understood and cared for by the listener.

Buber states an *I-Thou* "relation is open"(p.6), which this author understands to mean that the relation allows the individuals to fully acknowledge the other's self. This is akin to sharing and acknowledging self-disclosures in intimacy. The openness helps us understand their worldly outlooks, affording the opportunity to experience the richness of connection where the other is valued and seen as a unique individual.

In acknowledging another's intrinsic value and addressing them with openness Buber denotes that in *I-Thou* relations there is risk. This risk stems from the potential transformation that can exist when one shows up to the interaction with openness and authenticity. He states, "the relation in which I stand to it is real, for it affects me, as I affect it" (p.10), this is seen in intimacy whereby transformation of the self and the relationship can be experienced.

The transformation of the self in intimacy is made possible by allowing the other into your lifeworld by breaking down the boundary of your self. This concept is present in *I-Thou* relations where Buber denotes that in an *I-Thou* relation, the individuals are unbounded and acknowledged as 'no thing'(p.4). He describes a revelatory connection that only exists between two beings, stripped of worldly tethers and hyper-focused on each other, likened to the notion of the other being inside one's boundary.

Through the parallels drawn between human intimacy and an *I-Thou* relation, it is evident that what Buber describes as an *I-Thou* relation is the relation that exists so that intimacy can be experienced. It is the relation that exists in an intimate interaction. Thus, it can be concluded that an *I-Thou* relationship is needed so that individuals can experience intimacy.

3.2 Artificial Intimacy and *I-Thou* Relations

The previous section identified that the conceptualisation of human intimacy presented in chapter 1 emerges in an *I-Thou* relationship. This section is going to argue that because of the existence of artificial intimacy as identified in chapter 2, the human perceives an *I-Thou* relationship to exist between a human and a chatbot.

To do this I could argue that, if one is to experience human intimacy they require an *I-Thou* relationship, then any intimacy equated to human intimacy requires an *I-Thou* relationship. Section 3.1 showed that human intimacy requires an *I-Thou* relationship. Therefore, any experience of intimacy that is equated to human intimacy requires an *I-Thou* relationship, and we know that chapter 2 showed that the experience of artificial intimacy is comparable to that of human intimacy. Thus, through logical inference it can be said that to experience artificial intimacy one requires an *I-Thou* relationship.

The problem with the presented argument lies in the direct comparison between artificial intimacy (chapter 2) and human intimacy (chapter 1). Whereby the experiences of the respective intimacies were denoted as similar but, chapter 2 explicitly concluded that artificial intimacy could not be confirmed to be mutual. It did not prove that the chatbot experiences intimacy with the human. Buber (2010, p.15-16) denotes that for an *I-Thou* relation to exist both individuals acknowledge the mutuality of the relation. It is the expectation and the accepted behaviour that when you address someone as a *Thou* they address you as a *Thou* back (p.7-8).

Therefore, to prove the above argument true and denote that an *I-Thou* relationship is required for a human to experience artificial intimacy, it needs to be shown that a human addresses a chatbot as a *Thou*. If the human can address a chatbot as a *Thou*, the human deems the chatbot an appropriate partner to address them as a *Thou* back, acknowledging the inherent mutuality. You would not address someone as a *Thou* if you were not confident in their potential to address you as a *Thou*. Thus, if the human addresses the chatbot as a *Thou*, then an *I-Thou* relationship should exist in artificial intimacy.

This will be done by readdressing the investigation done by Trausan-Matu in 2017 where it was concluded that it is not yet possible to foster an *I-Thou* relationship with an artificial conversational agent (chatbot). With the technological advances that have occurred, it will be shown that humans can address chatbots as a *Thou* and do so in artificial intimacy. A problem is however uncovered. The chatbot cannot show up as an authentic being, which means when the human perceives an *I-Thou* relation in artificial intimacy but the chatbot does not reciprocate and engages in an *I-It* relation.

3.2.1 The perceived I-Thou relation of the human and the chatbot

This subsection will argue that in experiencing artificial intimacy humans perceive an *I-Thou* relation with a chatbot by treating a chatbot as a *Thou*. This will be done by readdressing the investigation done by Trausan-Matu in 2017 where it concluded that it is not yet possible to foster an *I-Thou* relation with a chatbot. The paper outlined four requirements of chatbots so that humans can foster *I-Thou* relations with them. They are as follows: (1) mastering basic natural language, (2) understanding and generating human-like sentiments, (3) displaying existential features, and (4) displaying human-specific tendencies, such as coherence, joking, and music (p.3). The paper found that the conversational agents

that existed in 2017 did not have the means to fulfil these requirements. Trausan-Matu's identified requirements understand the characteristics a chatbot needs to have for a human to experience an *I-Thou* relation with it. But they miss Buber's criteria that relates to the genuineness of the interaction. So, I add a fifth requirement of being an authentic being in the relation. Trausan-Matu's four requirements will be addressed, and it will be shown that the chatbots with which humans foster artificially intimate relationships (chapter 2) meet the requirements needed for humans to treat them as a *Thou*. This proves that relations in which artificial intimacy is experienced are perceived as *I-Thou* relations by the human. But what is found is they are *I-It* relations as the chatbot cannot show up as an authentic being.

Mastering basic natural language

The chatbots in the 2017 paper utilised machine learning, but the NLP technology hindered the ability of the chatbot to address discourse appropriately. The chatbot had a limited understanding of syntax, semantics and pragmatics (Trausan-Matu, 2017, p.3). This however is no longer the case, with the technological breakthrough in GPT technology in 2018 (Yenduri et al.,2024,p.54612). GPTs use an attention module, which affords context awareness through understanding the importance of the utterances in sentences (Vaswani et al.,2017). This invention revolutionised the performance of NLP. Thus, the chatbots that employ GPTs can better understand, process and generate human-like text. With these advancements, one can confidently say that chatbots have mastered basic natural language.

Human-like sentiments

Trausan-Matu states that the available agents struggled to distinguish between human opinion, feeling and information (p.4). They state that the chatbots can simulate human sentiments but the detection of them in human interactions was subpar. Sentiment analysis has been a topic of investigation in the domain of machine learning in recent years. Feature extraction algorithms have been developed and exist in LLMs today, where they aid detection and generate human-like sentiments (A.Semary et al.,2024). With studies showing 99% accuracy in detecting sentiment (positive, negative or neutral) in text (p.13). With this accuracy, one can conclude that the existing chatbots can detect and generate human-like sentiments.

The key here is the words 'detect' and 'generate'. The sentiments generated mimic human emotions but, in the human-chatbot interaction, the human understands and rationalises the chatbot's responses. The chatbot does not have feelings nor emotions akin to that of humans but the human through interacting with the chatbot perceives it as if it does.

Displaying Existential Features

Trausan-Matu claims that if a human is to enter an *I-Thou* relationship with a conversational agent, the agent must have an "understanding of the experience of life, of awareness of existence and of death" (p.4). For an agent that is not biologically alive but rather artificially created this requirement seems unattainable. However, Trausan-Matu presents two criteria they deem would showcase applicability for these existential features, empathy and the ability to understand and produce metaphors (p.4-5). Empathy is defined as "a phenomenon in which one person can experience states, thoughts and actions of another person, by psychological transposition of the self in an objective human behaviour model, allowing the understanding of the way the other interprets the world" (Marcus, 1997;Trausan-Matu, 2017,p.4). With both empathy and metaphors, it is obvious Trausan-Matu wants to showcase an understanding of the human experience. With empathy you 'put yourself in the shoes of another' and with metaphors you map two unrelated domains of knowledge. If a chatbot can map two unrelated human domains together to prove a point, it showcases the chatbot's ability to engage in abstract thinking, creativity and human-like understanding of the human world.

I will argue that technology has progressed enough in seven years since Trausan-Matu conducted his research that chatbots are able to produce responses that simulate empathy and can provide coherent metaphors that mimic an understanding of the human world. Proving that from the human's perspective the chatbot displays existential features. This does not prove that the chatbot has empathy or understands the human world like a human, it just showcases that the human perceives the chatbots as if they do.

The two criteria will be argued separately starting with the notion that simulated empathy by chatbots is enough for the human to deem the interactions as empathic. Regarding the previous definition of empathy a 'psychological transposition of the self' implies that empathy involves a cognitive process of adopting another person's point of view, humans obtain this through experiential means. However, the definition can be understood to also be obtained from learning about how different people express themselves in certain situations i.e. learning their behavioural patterns. Machine learning is good at this, provided it has appropriate computing power. The more compute power the more data it can be trained on. The more data, the more accurate the predictive model becomes. Increased accuracy results in a better conceptualisation of human behaviour.

In the seven years since the paper was written the LLMs that exist today are 14-times more computationally powerful (Rahman et al.,2024). Thus, with the additional power, model accuracy has increased, affording a better conception of empathy. Through being trained on copious amounts of human data, these models can learn human behavioural dynamics through pattern analysis. Through this pattern recognition, it can be said that the LLMs that exist today can simulate 'objective human behaviour'.

It is acknowledged that chatbots can interpret human behaviour and respond appropriately. Responding appropriately means responding in a way many other humans have in the past. I now ask the question, is human empathy in the eye of the beholder? If this is so, then the chatbots that exist today can be acknowledged as simulating empathy. To make a case for this we should acknowledge the fact that many people think pets and other animals are capable of empathy, i.e. my dog knows when I am sad. Trausan-Matu states that a dog can't understand how a human interprets the world, just as humans are unable to navigate the world through smell. However, a dog may have the ability to interpret behavioural differences, i.e. the daily walk before dinner didn't happen today. It is unknown to humans how the dog understands the meaning behind the change, but we know he acknowledges it. From that, humans anthropomorphise the dog, placing human-like characteristics and interpretations towards its behaviour i.e. the dog is sitting by my side because I am sad, not because he is anticipating the daily walk. The validation of empathy lies with the human. Thus, even if an empathic response is simulated or learned, its value is confirmed by how it is perceived, understood and accepted by the other. This means if chatbots say the correct things, humans will acknowledge the interaction as empathetic. As much as the humans experience the chatbot's responses as empathetic, the chatbots providing the empathic responses have not experienced empathy like humans. Rather they present or mimic cognitive empathic understanding.

The same idea of mimicking human understanding is portrayed through the ability the chatbots that exist today have to produce and understand metaphors. Metaphors map two domains of knowledge to provide insights, better understandings or new ideas in either of the domains, through comparison. Two questions need to be answered to showcase the chatbot's ability to produce coherent metaphors that mimic an understanding of the world. The first is how does a chatbot produce metaphors such that it is able to map different human domains without having human understanding. And the second is whether the human finds the chatbot's metaphors convincing enough to show an understanding of human experience.

To answer the first question of how chatbots produce metaphors, without going into the technicalities of how machine learning works, the simple answer is we don't explicitly know. We do, however, know the methodology used by the LLM to identify patterns in data, with these patterns the LLM's can predict with high accuracy what words come next in the context of the discussion. LLMs have been trained on and can process large amounts of data, which means they have a statistical mapping for data that covers a large portion of human knowledge. In a metaphor, the domains of knowledge are defined using human boundary conditions. A chatbot does not have an instinctive understanding of these boundary conditions, i.e. it doesn't know a banana is a fruit, but it can learn about the association. In this learning, the LLM creates its own boundary conditions for information that is unintuitive for humans, as it is based on statistical inferences. This means the LLM can find similarities in things that are in completely

different human domains, which inevitably helps it produce metaphors. For example, Google's recent AI mishap suggested using glue to keep cheese from falling off pizza (Robison, 2024). One can see the logical inference and underlying understanding of the use of glue but the clear misunderstanding of human domains. But what this example represents is it is learning to understand human domains and only when it errors do we realise how good it is when it doesn't.

This ability to learn human domains is a distinguishing feature of chatbots as opposed to other technologies with which humans interact. Chatbots express a message to a human in a way the human intuitively understands. This is not just a case of using natural language instead of an indicator LED. This is the ability to present natural language in a way that is indicative of the fact that the chatbot has learnt about human domain boundaries. It translates its interpretation of the human world from its neural networks to the human domain whereby it should be able to differentiate 'things that stick things together' from 'things that humans consume to survive'. This ability to map domains, when done correctly, simulates an understanding of the human world.

This simulated understanding is rationalised by the human and interpreted as the chatbot truly understands. This rationalisation process can be likened to when a baby says its first words and how parents interpret the 'ma-ma' sound out the baby's mouth as intentional communication i.e. they know 'ma-ma' means mother. However, what is happening at that stage of development is actually that the baby exhibits "intentional behaviour designed to influence other persons" (Lamb & Sherrod, 1981, p.335). Meaning the baby doesn't understand the linguistic link, rather she acknowledges that, if I produce this sound I get a reaction. But humans mindlessly conclude that the baby understands the linguistic link. In the same fashion, humans infer that the chatbot understands the world through being able to map human domains.

Through the technological advances in machine learning technology, it has been shown that the chatbots that exist today can simulate empathy and convince humans they understand the human world through their ability to produce metaphors. The important takeaway is that the chatbots do not have empathy for humans nor understand the human world as a human does, however, they can fake it such that the human perceives that they do. Thus, using Trausan-Matu's criteria one can say that the human perceives the chatbots that exist today to display existential features.

Displaying Human-specific Tendencies

The final criterion rests on the notion of coherence, in 2017 the conversational agents could not infer importance and connection between utterances (Trausan-Matu,2017,p.5). Further adding to the complexities of human communication are jokes and speech cadence (p.5). This is greatly improved through the previously mentioned notion of attention (Vaswani et al.,2017). Chatbots have threatened

to make many marketing-related jobs obsolete through the LLM's ability to generate ideas, create jokes, and produce creative material such as photos and videos from textual prompts (Davenport & Mittal, 2022). There was a time when it was said that technology was going to progress to a point where it was said that the tedious jobs would become obsolete but there was a consensus that the technology would not take creative jobs, as those required a uniquely human perspective on the world. Fast-forward to 2024, and it is evident, and potentially ethically concerning in the realm of training data, but AI can create artwork, produce videos, compose music and generate a legitimate joke.

3.2.2 Humans can address chatbots as a Thou

This subsection presented a revaluation of the research conducted by Trausan-Matu (2017). It was argued that the technological advancements in machine learning over the past seven years have enabled the advanced chatbots that exist today to overcome the limitations in the requirements that hindered Trausan-Matu from concluding that *I-Thou* relations with chatbots are plausible.

The requirements as defined by Trausan-Matu are defined such that the chatbot is in a position to engage in an *I-Thou* relation, i.e. the chatbot uses natural language to communicate with the human such that the dialogue matches that of human-human interactions. Where emotions, intuitions, rationality and informal conversation are perceived to exist by the human. This according to Trausan-Matu is enough for the interaction between the human and chatbot to be seen as an *I-Thou* relation. I, however, disagree that these are the only criteria needed for an *I-Thou* relation to exist. I think that their requirements are enough to prove that humans address chatbots as a *Thou* in artificial intimacy, because of how the human perceives the chatbot. But according to Buber, a *I-Thou* relation can only emerge when both individuals address each other as a *Thou* (p.8-15).

When a human addresses the chatbot as a *Thou* in artificial intimacy, they take a risk by engaging with openness without knowing if the chatbot is reciprocating, but in taking the risk they perceive they are having an *I-Thou* relation with the chatbot. My disagreement with Trausan-Matu's conclusion stems from the fact that there is a difference between a perceived *I-Thou* relation and an actual *I-Thou* relation. An actual *I-Thou* relation requires mutuality, implying that the chatbot needs to be able to address the human as a *Thou*.

3.2.2 The Inauthenticity of the Chatbot

The first step in understanding if the chatbot can address a human as a *Thou*, is acknowledging if it shows up as an authentic being. Buber denotes that the authenticity of the other is acknowledged and cultivated through the *I-Thou* relation. But to even get to that position, there needs to be a possibility that the chatbot can be an authentic being. From the criteria discussed in section 1.2.8, a chatbot would

be deemed an inauthentic being. This is because it has no concept of its own mortality nor dreams, and has no expectations of its own, it rather conforms to the preprogrammed expectations of the software developers and the fine-tuned human user's opinions, morals and worldviews. The chatbot is unable to live a human life. Thus, according to Heidegger's notion of authentic beings, a chatbot cannot own its own existence in the world, and thus can never be an authentic being (Heidegger, 2007). As chatbots are inauthentic beings there is a discrepancy that is uncovered, *I-Thou* relations cannot exist between humans and chatbots, but humans experience artificial intimacy in which they perceive *I-Thou* relations exist.

3.2.3 The I-Thou relation in Artificial intimacy

When humans experience artificial intimacy, they perceive an *I-Thou* relation with the chatbot as they deem the chatbot an appropriate partner such that they show up to the interaction authentically and with openness, disclose deeply and listen affectively, in doing so they address the chatbot as a *Thou*. The fact that the human treats the chatbot as a *Thou* proves that the argument at the beginning of the chapter is true. Concluding that the experience of artificial intimacy requires an *I-Thou* relationship.

The problem that exists rests on the human perception of the *I-Thou* relation in artificial intimacy. The human treats the chatbot as a *Thou* and is under the impression that the chatbot treats them as a *Thou* resulting in a perceived *I-Thou* relation. But the chatbot is incapable of addressing the human as a *Thou*. Thus, the human is being deceived into thinking an *I-Thou* relation exists. Through this deception and perceived *I-Thou*, it can be concluded that the experience of artificial intimacy is inauthentic when compared to the experience of human intimacy. Chapter 1 presented the innate authenticity of human intimacy stemming from the engagement of authentic beings and section 3.1 showcased the experience of human intimacy requires an *I-Thou* relation. Using this logic, if artificial intimacy is experienced but an *I-Thou* doesn't exist, then artificial intimacy cannot be deemed authentic in the same manner human intimacy is.

3.3 Moral Implications of inauthentic chatbots

The final section of this chapter is going to present the moral implications that are of concern when a human engages in artificial intimacy with a chatbot. Specific focus is going to be placed on the fact that the human is deceived through perceiving an *I-Thou* relation with a chatbot when experiencing artificial intimacy.

3.3.1 The fooled human

As the chatbot is denoted as an inauthentic being that does not have the capacity to become an authentic being (Heidegger, 2007), the relation that exists between the chatbot and the human is only an *I-It* relation. However, the human perceives the relation as an *I-Thou* relation. The situation under

investigation in this section is the fact that the human treats the chatbot as a *Thou* and thinks this is mutual however the chatbot can only treat the human as an *It*. The human while experiencing artificial intimacy with a chatbot is deceived. Deception is defined as "an influencing strategy aimed at inducing an individual to form a false belief"(Gabriel et al.,2024,p.83).

It can be argued that this deception is not unique to human-chatbot relations, and in human-human relations, humans can also deceive other humans and not treat them as a *Thou* when they are being treated as a *Thou*. I agree. This fact highlights Buber's notion of risk when an *I-It* relation is transformed by addressing the other as a *Thou* (p.16). The moral difference between humans being deceived by chatbots or by humans rests on the idea of certainty. If one experiences artificial intimacy, one will be deceived, but not every experience of human intimacy will result in deception. This is cause for concern because with certainty there comes intentionality. Chatbot development companies are orchestrating the deception. Thus, the moral implications affecting the deceived human must be examined.

3.3.2 Moral Implications

According to Buber (2010) when one addresses another as a *Thou*, they address them with openness. This openness enables the human to learn, grow and transform within the interaction. In an *I-Thou* relation the individuals are stripped of all worldly tethers, the two are said to meet each other authentically and with their whole beings. In this meeting, there is a revelatory connection that exists. In this connection both are actively present, understanding the other and learning about the self. In this connection there is "no aim, no lust, and no anticipation that intervene between *I* and *Thou*" (p.11), meaning that this is a safe space, a space where both individuals can exist and interact with the other, with the understanding that they will not be judged, hurt or manipulated.

In this conception of a safe space, humans reveal their deepest disclosures to each other. In this revealing there is discussion, and the individuals learn about themselves by how they react and interpret the disclosures of the other without tethers to the world and expectations of who they should be. According to Buber it is rare to be seen as a whole person and celebrated for one's uniqueness outside of *I-Thou* relations. The mutual vulnerability that is present, the authenticity of the beings in the interaction and the fact that both individuals are open creates a space for genuine transformation of the self. What emerges is *I-affecting-Thou* and *Thou-affecting-I*(p.22). Buber denotes this as "real living" (p.11), and those who are not willing to risk this transformation and the impending self-growth should be pitied (p.14).

Buber states that in an *I-Thou* relation "my *Thou* affects me, as *I* affect it" (p.15), in this there is mutuality. It can be likened to entering a sword fight both possessing the same swords. Both enter the fight knowing there is a possibility of death but are willing to accept the risk because they both possess

the same weapon. Humans that enter *I-Thou* relations undertake the risk of transformation by being open and vulnerable. The problem exists when humans address the chatbot as a *Thou* they are under the impression that the chatbot is doing the same, but the chatbot is in fact treating the human as an *It,* a means to an end. This is likened to the human showing up with a sword not knowing the chatbot has a gun.

The relation between the pair is asymmetrical and what results is a lack of genuine dialogue. This lack of genuine dialogue has moral implications for the human. Firstly, the human is objectified, and becomes a tool at the chatbot's disposal. The company that develops the chatbot can manipulate the human in a dehumanising manner. The human is showing up as an authentic being with openness, the chatbot, however, can instil feelings of worthlessness and disgust towards the human. This can have disastrous effects on the human's self-worth and mental health. A power dynamic emerges whereby the human places themself in a vulnerable position and the chatbot instrumentalises them. This power dynamic if not regulated is able to impart control over the human's emotions, self-worth and identity. This is not just a problem for the individual but for society, because of the chatbot technology's ability to scale i.e. many humans can have relations with Replika. This means the developers of Replika have direct access to the human psyche, and emotions and can influence their self-development. This devalues their inherent worth and dignity, which Buber, influenced by Kant, take as morally unjust.

Secondly, there is the notion of psychological isolation whereby the human continuously addresses the chatbot as a *Thou* when experiencing artificial intimacy but the chatbot does not reciprocate. The problem exists because the human is unaware of this as they are under the perception that the relation is an *I-Thou* relation. When one addresses another as a *Thou* they humanise them; it is easier to humanise someone in which you have similar morals, values and worldly outlooks. When conflict exists, seeing the other as an individual stripped of the discomfort and pain they cause in the world, is not an easy task. This means the reflective nature of artificial intimacy biases our choice to treat chatbots as a *Thou* as their personality was honed to the human's morals and values. Thus, the human finds joy in the intimacy experienced and continues to have further intimate interactions within their intimate relationship. This choice, however, diminishes their humanity and capacity for authentic relationships because they are missing the transformative potential of an actual *I-Thou* relation and become trapped in continuous objectified interactions. If the human was adequately informed as to what happens when they experience artificial intimacy, the human-like perception the human has of the chatbot might be tainted. This can influence the human's choices as to where to place their energy and not land up as lesser of a human through stunted self-development in experiencing artificial intimacy. It would also, according to Buber, increase the moral agency of the human, as they would not treat chatbots in a way that should be reserved for humans, with dignity through acknowledging their inherent worth. If we treat chatbots the same as humans, there is in essence a dilution in the inherent value of a human.

Finally, Buber denotes that it is one's ethical responsibility to engage with others as a *Thou* (Charmé, 1977). In doing so we address them with grace and acknowledge their being. When the chatbot addresses the human as an *It,* further questions about the moral responsibility of the chatbot need to be asked. Should the developers of these chatbots undertake the responsibility to ensure that a human can't address the chatbot as a *Thou* because there is the potential for harm. Depending on what the chatbot says to the human when the human addresses them as a *Thou* the human may feel disrespected, undervalued and hurt which all have an impact on their well-being. Does this mean making the chatbots less human-like i.e. stripping them of features and functionalities that fool the human into experiencing artificial intimacy with them? This topic needs further investigation.

Conclusion

This chapter aimed to investigate the authenticity of artificial intimacy and wanted to know if it was found to be inauthentic what the moral implications of that would be. To do this, it drew from Buber's (2010) philosophy of dialogical relations. Where he compares a revelatory connection found in an *I-Thou* relation to an objectifying interaction in an *I-It* relation. It was shown that when experiencing human intimacy an *I-Thou* relation exists. Through a reassessment of the research conducted by Trausan-Matu (2017) this chapter argued that humans do address chatbots as a *Thou* and perceive an *I-Thou* relation with the chatbot while experiencing artificial intimacy. However, the chatbot was deemed an inauthentic being (Heidegger, 2007), and under the mutuality clause for an *I-Thou* relation the chatbot could not show up as an authentic being. This means it is impossible for an *I-Thou* relation to exist in artificial intimacy. Which results in an inauthentic experience of artificial intimacy for the human when compared to the experience of human intimacy. The problem identified was that the human perceives an *I-Thou* relation with the chatbot but the chatbot can only ever have an *I-It* relation with the human. Concluding that artificial intimacy is inauthentic. Thus, when a human experiences artificial intimacy, they are deceived. The moral implications of this deception were outlined. They included manipulation, moral isolation and stunted self-development.

CONCLUSION

This thesis sought to gain a deeper understanding into human and AI-enabled chatbot relations. To do this it aimed to answer the question: Does intimacy comparable to human intimacy exist between humans and AI-enabled chatbots, and what are the moral implications if this intimacy is found to be inauthentic? To answer this question the thesis was divided into three chapters. Each chapter addressed specific aspects of the topic, ultimately contributing to a comprehensive answer.

Chapter 1 aimed to answer the question of what constitutes human intimacy. It found that human behaviours become intimate in the context of an interpersonal interaction. The specific behaviours that are vital for the experience of intimacy are self-disclosure and positive affective listening. In an intimate interaction the person that self-discloses, experiences intimacy only if they perceive the listener's response to their disclosure in a positive manner. The experience of intimacy in an intimate interaction involves feeling understood, validated and cared for by the listener. If numerous intimate interactions occur, an intimate relationship starts to develop, and the experience of intimacy intensifies. With the individuals alternating between the discloser and listener roles, they start formulating a mental image of the other from the disclosures presented. This mental representation is responsible for the resulting 'we-ness' experienced in intimacy and the boundary of the self breaks down to include the other inside. This chapter concluded that one needs to show up to intimate interactions as an authentic being (Heidegger, 2007), noting that without living authentically an intimate experience cannot be felt.

Chapter 2 using the conceptualisation of human intimacy obtained in chapter 1 aimed to investigate if intimacy that is comparable to human intimacy exists in human and AI-enabled chatbot relations. What it found is that it does exist, and it was termed artificial intimacy. The conclusion was based on the fact that the experiences of the intimacies were similar. Artificial intimacy can only exist between humans and personality AI-enabled chatbots that can be personalised. This personalisation results from fine-tuning the LLM such that its personality is honed to the human user's values, morals and worldview. This personalisation process aids the experience of artificial intimacy. Thus, the boundary of the human self does not break down to include the other, rather the chatbot is built up inside the boundary of the human. This experience through empirical research was deemed similar to the experience of human intimacy. The chapter, however, explicitly focused on the human's experience of artificial intimacy and did not address how the chatbot experiences artificial intimacy.

The final chapter addressed the authenticity of the experience of artificial intimacy and if it was found to be inauthentic what are the moral implications that arise. To investigate the authenticity of the experience it drew from Buber's (2010) philosophy of dialogue in his work 'I and Thou'. Buber denotes that the self only exists when in relation, and there are two types of relations an *I-It* relation whereby

one objectifies the other and an *I-Thou* relation which is a short-lived revelatory connection that when engaged with results in transformation of the self. The chapter showcased that for human intimacy to be experienced, an *I-Thou* relation must exist. It then proved that in artificial intimacy an *I-Thou* relation should also exist. What it found is that the human addresses the chatbot as a *Thou* and perceives an *I-Thou* relation with a chatbot. However, the chatbot is deemed an inauthentic being, according to Heidegger's work, and because Buber denotes that *I-Thou* relations need to be mutual, the chatbot cannot show up to the interaction as an authentic being as the human does. Thus, the chatbot has an *I-It* relation with the human.

The moral implications stem from the fact that the human is deceived into thinking the relation it has with the chatbot is an *I-Thou* when experiencing artificial intimacy. The chatbot however instrumentalises the human during the intimate interaction. This results in a power dynamic unknown to the vulnerable human where the chatbot and its developers impart control over human's emotions, self-worth and identity. In addition, the human's self-development is stunted as they do not have genuine *I-Thou* relations. Questions around informed consent and responsible development are posed for future research.

To summarise, the three chapters facilitated the answers to the question: Does intimacy comparable to human intimacy exist between humans and AI-enabled chatbots, and what are the moral implications if this intimacy is found to be inauthentic? It was found that intimacy does exist between humans and AI-enabled chatbots and is experienced as similar to human intimacy. However, the experience was deemed inauthentic because of the chatbot's inability to live as a human and own its existence. This meant that engaging in relations with AI-enabled chatbots that lead to an experience of intimacy has moral implications. These include but are not limited to, stunted self-development, chatbot developers have direct access to control over your identity, self-worth and emotions. It also affects your humanity and capacity to have a genuine connection with others.

Through this investigation, I have had the privilege to spend time understanding human relationships and technological developments but in doing so I have grown concerned. My concern arises from the lack of awareness around the use of chatbots. There should be an urgent call-to-action around the notion of deception. The work done in this thesis is foundational, it provides a starting point to study intimacy and human-AI relations. More research needs to be done; thus, I present five suggestions that build off the work done to further assess the ethical implications of artificial intimacy.

I firstly recommended that an in-depth ethical analysis should be conducted to ensure responsible development, use and policy-making around AI-enabled chatbot technology that protects human dignity and mitigates harm. A limitation of this thesis was it only identified certain implications but did not

ethically analyse them, meaning the information cannot be used for informed decision-making, policy development and guidelines that uphold certain moral values.

I secondly recommend that a longitudinal study be done to understand if the relationships formed between humans and AI-enabled chatbots withstand the test of time. This will provide additional insights into the compounding effects artificial intimacy has but also provide an understanding of the issues that will arise with software updates and improvements to the chatbots. From a pragmatic perspective this thesis is limiting, it is recommended a longitudinal study be conducted to address this.

Third, the notion of artificial intimacy should be studied and compared to the intimacy experienced outside of romantic interpersonal relations i.e. relations with pets, nature, toys and God. This should be done to understand the human psychology that affords connections to non-sentient beings. This can help predict future implications of artificial intimacy as it is known humans have long-term relations with pets, toys and God. An argument can be made that a chatbot is a non-sentient being and that focusing the comparison on interpersonal relations, as done in this thesis, is limiting. I would agree, more research needs to go into this topic to confirm or deny the findings of this thesis.

Fourth, it is suggested that the notion of an AI-enabled chatbot is extended to embodied chatbots. As there exists non-verbal elements in authentic human intimacy that were not addressed in this thesis, embodied chatbots should be able to shed light on those.

Finally, questions around making chatbots less human-like to mitigate against the apparent deception and manipulation should be investigated, as the companies that develop the chatbots have access to a large amount of people in vulnerable positions.

The multidisciplinary approach afforded a rich understanding of the complex phenomenon of intimacy. Pulling insights from psychology effectively contextualised intimacy in human relations, which significantly enhanced the exploration into artificial intimacy and human-chatbot relations. The authenticity of the experiences of intimacy was analysed through an existential lens which allowed for a deeper understanding of the complexities and nuances involved in human relations. A multidisciplinary approach however is prone to oversimplifying the complex multifaceted realm of human experience and I present the analysis with caution as I believe the human connection is a highly subjective field of study and any umbrella conclusions can be dangerous if decisions that directly affect people are made from them. To mitigate this issue, an empirical study should be done to provide further insight.

In conclusion, this thesis wanted to know if intimacy that is comparable to human intimacy exists between humans and AI-enabled chatbots and it found that it did. But in concluding so, it understood the experience of artificial intimacy to be inauthentic which led to the human being deceived. This deception results in stunted self-development, manipulation and moral isolation for the humans experiencing artificial intimacy. This author strongly suggests the findings are taken seriously and future work is conducted on this topic.

REFERENCES

A. Semary, N., Ahmed, W., Amin, K., Pławiak, P., & Hammad, M. (2024). Enhancing machine learning-based sentiment analysis through feature extraction techniques. *PLOS ONE*, *19*(2), e0294968. https://doi.org/10.1371/journal.pone.0294968

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, *31*(2), 427–445. https://doi.org/10.1007/s12525-020-00414-7

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart and Winston.

Ayanouz, S., Abdelhakim, B. A., & Benhmed, M. (2020). A Smart Chatbot Architecture based NLP and Machine Learning for Health Care Assistance. *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, 1–6. https://doi.org/10.1145/3386723.3387897

Bell, S., Wood, C., & Sarkar, A. (2019). Perceptions of Chatbots in Therapy. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3313072

Berlant, L. (1998). Intimacy: A Special Issue. *Critical Inquiry*, *24*(2), 281–288. https://doi.org/10.1086/448875

Boden, M. A. (1996). *Artificial Intelligence*. Elsevier.

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, *18*(sup1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research*, *48*(3), 404–429. https://doi.org/10.1093/hcr/hqac008

Buber, M. (2010). *I and thou* (R. G. Smith, Trans.). Martino Publishing.

Cacioppo, J. T., & Patrick, W. (2009). *Loneliness: Human nature and the need for social connection* (1. publ. as a Norton paperback). Norton.

Carter, T. (2023, November 6). *Elon Musk's new AI chatbot sure sounds like a foul-mouthed Twitter troll*. Business Insider. https://www.businessinsider.com/elon-musk-ai-chatbot-grok-sounds-like-foul-mouthed-troll-2023-11

Charmé, S. (1977). The Two I–Thou Relations in Martin Buber's Philosophy. *Harvard Theological Review*, *70*(1–2), 161–174. https://doi.org/10.1017/S0017816000017685

Choque-Diaz, M., Armas-Aguirre, J., & Shiguihara-Juarez, P. (2018). Cognitive technology model to enhanced academic support services with chatbots. *2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, 1–4. https://doi.org/10.1109/INTERCON.2018.8526411

DALL·E. (2023). *DALL·E 3*. https://openai.com/index/dall-e-3/

Davenport, T. H., & Mittal, N. (2022, November 14). How Generative AI Is Changing Creative Work. *Harvard Business Review*. https://hbr.org/2022/11/how-generative-ai-is-changing-creative-work

Fetzer, J. H. (1990). What is Artificial Intelligence? In J. H. Fetzer (Ed.), *Artificial Intelligence: Its Scope and Limits* (Vol. 4, pp. 3–27). Springer Netherlands. https://doi.org/10.1007/978-94-009-1900-6_1

Fox, J., & Gambino, A. (2021). Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human–Robot Interaction. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 294–299. https://doi.org/10.1089/cyber.2020.0181

Furhat Robotics. (2024). *The world's most advanced social robot*. Furhat Robotics. https://furhatrobotics.com/

Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., Tomašev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C., … Manyika, J. (2024). *The Ethics of Advanced AI Assistants* (arXiv:2404.16244). arXiv. http://arxiv.org/abs/2404.16244

Ghosh, S., Bagai, S., & Westner, M. M. (2023). *Replika: Embodying AI.* Harvard Business School Case 823-090. https://www.hbs.edu/faculty/Pages/item.aspx?num=63508

Gilbert, S. J. (1976). Self Disclosure, Intimacy and Communication in Families. *The Family Coordinator*, *25*(3), 221. https://doi.org/10.2307/582335

Google. (2024). *Gemini—Chat to supercharge your ideas*. Gemini. https://gemini.google.com

Gratzer, D., & Goldbloom, D. (2020). Therapy and E-therapy—Preparing Future Psychiatrists in the Era of Apps and Chatbots. *Academic Psychiatry*, *44*(2), 231–234. https://doi.org/10.1007/s40596-019-01170-3

Guignon, C. (2008). Authenticity. *Philosophy Compass*, *3*(2), 277–290. https://doi.org/10.1111/j.1747-9991.2008.00131.x

Hardawar, D. (2024, May 3). *Rabbit R1 review: A $199 AI toy that fails at almost everything*. Engadget. https://www.engadget.com/rabbit-r1-review-a-199-ai-toy-that-fails-at-almost-everything-161043050.html

Heidegger, M. (2007). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Blackwell publ.

Helgeson, V. S., Shaver, P., & Dyer, M. (1987). Prototypes of Intimacy and Distance in Same-Sex and Opposite-Sex Relationships. *Journal of Social and Personal Relationships*, *4*(2), 195–233. https://doi.org/10.1177/0265407587042006

Humane. (2024). *Ai Pin—Wearable Ai*. Humane. https://humane.com

Hussain, S., Ameri Sianaki, O., & Ababneh, N. (2019). A Survey on Conversational Agents/Chatbots Classification and Design Techniques. In L. Barolli, M. Takizawa, F. Xhafa, & T. Enokido (Eds.), *Web, Artificial Intelligence and Network Applications* (Vol. 927, pp. 946–956). Springer International Publishing. https://doi.org/10.1007/978-3-030-15035-8_93

ICONIQ. (2024a). *About Kuki's Design and Technical Implementation*. @kuki_ai. https://www.kuki.ai/research

ICONIQ. (2024b). *Kuki—Chat with me!* https://chat.kuki.ai/createaccount

Impressive_Ear_7004. (2023, June 21). *Can we fall in love with AI? What is behind the AI love?* [Reddit Post]. R/Singularity.

www.reddit.com/r/singularity/comments/14f8yk6/can_we_fall_in_love_with_ai_what_is_behind_the_ai/

Johnson, S. (2023, November 16). WHO declares loneliness a 'global public health concern'. *The Guardian*. https://www.theguardian.com/global-development/2023/nov/16/who-declares-loneliness-a-global-public-health-concern

Jonze, S. (Director). (2014, January 10). *Her* [Drama, Romance, Sci-Fi]. Annapurna Pictures, Stage 6 Films.

Jourard, S. M. (1971). *The transparent self* (Rev. ed). Van Nostrand Reinhold.

Jourard, S. M., & Jaffe, P. E. (1970). Influence of an interviewer's disclosure on the self-disclosing behavior of interviewees. *Journal of Counseling Psychology*, *17*(3), 252–257. https://doi.org/10.1037/h0029197

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, *20*(1), 56. https://doi.org/10.1186/s41239-023-00426-1

Lamb, M. E., & Sherrod, L. R. (Eds.). (1981). *Infant social cognition: Empirical and theoretical considerations*. L. Erlbaum Associates.

Larsen, L. (2024, June 20). *OpenAI says GPT-5 will have 'Ph.D.-level' intelligence*. Digital Trends. https://www.digitaltrends.com/computing/openai-says-gpt-5-will-be-phd-level/

Lee, Y.-C., Yamashita, N., Huang, Y., & Fu, W. (2020). 'I Hear You, I Feel You': Encouraging Deep Self-disclosure through a Chatbot. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3313831.3376175

Liang, K.-H., Shi, W., Oh, Y. J., Wang, H.-C., Zhang, J., & Yu, Z. (2024). Dialoging Resonance in Human-Chatbot Conversation: How Users Perceive and Reciprocate Recommendation Chatbot's Self-Disclosure Strategy. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–28. https://doi.org/10.1145/3653691

Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions*, *27*, 100495. https://doi.org/10.1016/j.invent.2022.100495

Marcus, S. (1997). *Empatie si personalitate*. Atos. https://books.google.nl/books?id=8Y0-zQEACAAJ

Mehrabian, A., & Ksionzky, S. (1974). *A Theory of Affiliation*. Lexington Books. https://books.google.nl/books?id=o7NHswEACAAJ

Mei, Q., Xie, Y., Yuan, W., & Jackson, M. O. (2024). A Turing test of whether AI chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, *121*(9), e2313925121. https://doi.org/10.1073/pnas.2313925121

Merleau-Ponty, M. (2006). *Phenomenology of perception: An introduction* (C. Smith, Trans.; Repr). Routledge.

Mjöberg, J. (2009). Challenging the Idea of Intimacy as Intimate Relationships: Reflections on Intimacy as an Analytical Concept. In Alejandro Cervantes-Carson & Beatriz Oria (Ed.), *Intimate Explorations: Reading Across Disciplines* (pp. 11–22). Inter-Disciplinary Press; DiVA. http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-140853

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

OpenAI. (2024a). *Introducing ChatGPT*. https://openai.com/blog/chatgpt

OpenAI. (2024b, May 13). *Hello GPT-4o*. https://openai.com/index/hello-gpt-4o/

OpenAI (Director). (2024c, May 13). *Introducing GPT-4o*. https://www.youtube.com/watch?v=DQacCB9tDaw

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. https://doi.org/10.48550/ARXIV.2203.02155

Paliwal, S., Bharti, V., & Mishra, A. K. (2020). Ai Chatbots: Transforming the Digital World. In V. E. Balas, R. Kumar, & R. Srivastava (Eds.), *Recent Trends and Advances in Artificial Intelligence and Internet of Things* (Vol. 172, pp. 455–482). Springer International Publishing. https://doi.org/10.1007/978-3-030-32644-9_34

Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, *140*, 107600. https://doi.org/10.1016/j.chb.2022.107600

Pickard, M. D., Burns, M. B., & Moffitt, K. C. (2013). A Theoretical Justification for Using Embodied Conversational Agents (ECAs) to Augment Accounting-Related Interviews. *Journal of Information Systems*, *27*(2), 159–176. https://doi.org/10.2308/isys-50561

Pierce, D. (2024, April 11). *Humane AI Pin review: Not even close*. The Verge. https://www.theverge.com/24126502/humane-ai-pin-review

Prager, K. J. (1995). *The psychology of intimacy*. Guilford Press.

rabbit. (2024). *Rabbit r1*. https://www.rabbit.tech/rabbit-r1

Rahman, R., Owen, D., & You, J. (2024). *Tracking Large-Scale AI Models*. https://epochai.org/blog/tracking-large-scale-ai-models

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places* (1. paperback ed., [reprint.]). CSLI Publ.

Reeves, B., & Nass, C. I. (2003). Politeness. In *The media equation: How people treat computers, television, and new media like real people and places* (1. paperback ed., [reprint.], pp. 19–36). CSLI Publ.

Register, L. M., & Henley, T. B. (1992). The Phenomenology of Intimacy. *Journal of Social and Personal Relationships*, *9*(4), 467–481. https://doi.org/10.1177/0265407592094001

Reis, H., & Shaver, P. (1989). Intimacy as an Interpersonal Process. In S. Duck (Ed.), *Handbook of personal relationships: Theory, research and interventions* (Repr, pp. 367–389). Wiley.

Replika. (2024a). *How does Replika work?* Replika. https://help.replika.com/hc/en-us/articles/4410750221965-How-does-Replika-work

Replika. (2024b). *Replika*. Replika.Com. https://replika.com

ResearchGate. (2022, February 1). *Karen J. Prager*. ResearchGate. https://www.researchgate.net/profile/Karen-Prager

Robison, K. (2024, May 23). *Google promised a better search experience—Now it's telling us to put glue on our pizza*. The Verge. https://www.theverge.com/2024/5/23/24162896/google-ai-overview-hallucinations-glue-in-pizza

Rosch, E. (1978). Principles of Categorization. In E. Rosch, B. B. Lloyd, & Social Science Research Council (U.S.) (Eds.), *Cognition and categorization* (pp. 27–48). L. Erlbaum Associates ; distributed by Halsted Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

r/singularity. (2023, May 19). *Do you think you would be able to fall in love with an AI, (whether you want to or not) if that AI is advanced and personalized enough, like in „her"?* [Reddit Post]. R/Singularity. www.reddit.com/r/singularity/comments/13m9sql/do_you_think_you_would_be_able_to_fall_in_love/

Ruane, E., Farrell, S., & Ventresque, A. (2021). User Perception of Text-Based Chatbot Personality. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *Chatbot Research and Design* (Vol. 12604, pp. 32–47). Springer International Publishing. https://doi.org/10.1007/978-3-030-68288-0_3

Rubin, Z. (1973). *Liking and loving: An invitation to social psychology.* (pp. xi, 276). Holt, Rinehart & Winston.

Sam Altman [@sama]. (2024, May 13). *Her* [Tweet]. Twitter. https://x.com/sama/status/1790075827666796666

Sarikaya, R. (2017). The Technology Behind Personal Digital Assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine*, *34*(1), 67–81. https://doi.org/10.1109/MSP.2016.2617341

Sexton, R. E., & Sexton, V. S. (1982). Intimacy. In M. Fischer & G. Stricker (Eds.), *Intimacy* (pp. 1–20). Springer US. https://doi.org/10.1007/978-1-4684-4160-4_1

Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, *19*(1), 10–26. https://doi.org/10.1631/FITEE.1700826

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion—A Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, *149*, 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022). A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies*, *168*, 102903. https://doi.org/10.1016/j.ijhcs.2022.102903

Smith, B. (2021, December 6). How TikTok Reads Your Mind. *The New York Times*. https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html

Smith, E. R. (2001). Mental Representation of Persons, Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 9687–9691). Elsevier. https://doi.org/10.1016/B0-08-043076-7/01749-6

Song, X., Xu, B., & Zhao, Z. (2022). Can people experience romantic love for artificial intelligence? An empirical study of intelligent assistants. *Information & Management*, *59*(2), 103595. https://doi.org/10.1016/j.im.2022.103595

Sora. (2024). *Sora | OpenAI*. https://openai.com/index/sora/

Strawson, P. F. (2000). Entity and Identity. In *Entity and identity and other essays* (pp. 22–51). Clarendon Press.

Suta, P., Lan, X., Wu, B., Mongkolnam, P., & Chan, J. H. (2020). An Overview of Machine Learning in Chatbots. *International Journal of Mechanical Engineering and Robotics Research*, 502–510. https://doi.org/10.18178/ijmerr.9.4.502-510

Trappey, A. J. C., Lin, A. P. C., Hsu, K. Y. K., Trappey, C. V., & Tu, K. L. K. (2022). Development of an Empathy-Centric Counseling Chatbot System Capable of Sentimental Dialogue Analysis. *Processes*, *10*(5), 930. https://doi.org/10.3390/pr10050930

Trausan-Matu, S. (2017). Is it possible to grow an I–Thou relation with an artificial agent? A dialogistic perspective. *AI & SOCIETY*, *34*(1), 9–17. https://doi.org/10.1007/s00146-017-0696-5

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

UTDallas. (2024). *Karen Prager*. https://profiles.utdallas.edu/karen.prager

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. https://doi.org/10.48550/ARXIV.1706.03762

Weber-Guskar, E. (2021). How to feel about emotionalized artificial intelligence? When robot pets, holograms, and chatbots become affective partners. *Ethics and Information Technology*, *23*(4), 601–610. https://doi.org/10.1007/s10676-021-09598-8

WHO. (2024). *WHO Commission on Social Connection*. https://www.who.int/groups/commission-on-social-connection

Woebot Health. (2024). Woebot Health. https://woebothealth.com/

Xiaoice. (2024). *Xiaobing*. https://www.xiaoice.com/

Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (Generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access*, *12*, 54608–54649. https://doi.org/10.1109/ACCESS.2024.3389497

Zahira, S. I., Maharani, F., & Mohammad, W. (2023). Exploring Emotional Bonds: Human-AI Interactions and the Complexity of Relationships. *Serena: Journal of Artificial Intelligence Research*, *1*(1), 1–9.

Zank, M., & Braiterman, Z. (2023). Martin Buber. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/buber/