Investigating human beings' experience with Artificial Intelligent social robots and/or digital avatars in an interactional context; what is happening and how it informs the future

Brigitte Juliana Viljoen

2023

School of Clinical Science, Department of Psychotherapy and Counselling

A thesis submitted to
Auckland University of Technology
in fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

Faculty of Health and Environmental Sciences – Te Ara Hauora A Pūtaiao

Abstract

To date, research conducted in human-robot interaction has primarily focused on how to make robots more appealing and comfortable for humans to interact with (Dautenhahn, 2007; Sciutti et al., 2018). An area that has received less attention, is research into humans' experiences when interacting with an Artificial Intelligence (AI) social robot and/or digital avatar. Utilising Charmaz' (2003, 2006, 2014) constructivist grounded theory methodology, this study investigated how human beings experienced AI social robots and/ or digital avatars in an interactional context. In particular, the research explored the affective dimensions of these interactions; including ways of relating and connection. In addition, the research considered the potential impact on human beings applying this technology in various fields of human communication.

A major finding of the study shows that human beings interacting with an AI social robot and/or digital avatar can, when immersed in a conversational flow state, develop unconscious feelings towards these AI social systems, as if they are real people. In other words, an 'Other' was constructed by the human end-user. This experience can feel as real as human-to-human interaction. A key insight that emerged is that human beings can feel a deep-seated need to connect to others which, by being in the moment, is not consciously apprehended until after the fact. When these types of humanlike interactions occur, a trust develops towards this type of AI social system, leading to a strong sense of the AI robot and/or avatar 'as if' it was a real person.

The research contributes to a deeper understanding of the processes underlying human interaction with AI social systems, such as AI social robots and/or digital avatars. The findings can assist technologists in their efforts to develop and market their products ethically; and direct their efforts towards beneficial uses rather than those which exploit humans. In addition, the findings may be useful for human end-users and third parties developing greater awareness of the impacts of such technology on human beings.

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signed:

Name: Brigitte Viljoen

Acknowledgements

This PhD would not have been possible without the support of many people.

First, I would like to acknowledge the participants of this research. My sincerest gratitude and appreciation to the people who generously shared their time and trusted me with their experiences in order to make this all possible. Your contribution to this research is appreciated and very much valued. I would also like to thank the number of people that took the time to meet with me to talk about my research topic and share their expertise, which I found invaluable. A special thank you is for Dr Elizabeth Broadbent who, at various stages in my PhD journey, gave me sound advice, support and encouragement. Thank you.

I am very grateful for my supervisors, at the beginning of my PhD process; Professor Keith Tudor and Dr Brian Rodgers were willing to start me on my PhD journey, with Dr Brian Rodgers becoming my primary supervisor subsequently, followed by Dr Elizabeth Day graciously agreeing to be my primary supervisor when Dr Brian Rodgers left AUT. I would also like to thank Dr Mahsa McCauley for stepping in later in the PhD process to share and provide supervision on her expertise in artificial intelligence. Additionally, a special thank you to Dr Kim Ward for sharing her expertise and supervision on ground theory methodology and method at the beginning part of my PhD. I have appreciated all of your guidance, encouragement, support, time, and being challenged from our discussions, as well as your experience, expertise, and suggestions. Thank you.

I would like to acknowledge the experience, support, and encouragement that I have received from the regular AUT University Grounded Theory Group which has been invaluable in my PhD journey. Not only have I learned and developed so much from discussions, questions, and sharing of each other's research on grounded theory, the group has provided compassion and support for me throughout my PhD process overall. I am thankful I had a secure and encouraging space to develop as a researcher over the years, and to build my confidence within grounded theory research. Thank you to the grounded theory group.

I was thankful to have received a Doctoral Fee Scholarship (Faculty of Health and Environmental Sciences) for 3-years from Auckland University of Technology, without which I would not have been able to do this PhD. Thank you to Shoba Nayar for her help, efficiency, and expertise with transcribing, proof reading, and formatting the final thesis.

I would like to acknowledge that this PhD could not have been completed without the ongoing support, patience, and encouragement of a number of family, friends, and colleagues along the way. A few need a special mention. Thank you to my friends, Dr Charles and Ria Sanders, Reneé Olivier, Melanie Kozak, and many others. I appreciate your understanding, support, and encouragement; and your ongoing friendship and help when I needed it. Of course, a special thank you to my husband, Gabriel, for his love, understanding, and support from the beginning until the end of the PhD. Thank you for taking on more of the household duties to give me the time I needed to complete this thesis. Thank you also to my son, Onree, for giving me the space to work – even when you wanted my attention – and for being patient and supportive with me. Thank you to my mother-in-law, Hester, for your ongoing support, love, and respect from overseas. Also, a special thank you to my late parents who I know would have been supportive and proud of me completing this PhD. I could not have done this without my family and friends, who have encouraged and moved with me throughout this whole PhD journey.

Ethical approval for this research was granted by Auckland University of Technology Ethics Committee on 24 October 2019, reference number 19/390.

Technological change is social through-and-through.
Take away the… structures that support technological change of a particular sort,
and it ceases to seem 'natural' – indeed it ceases altogether.
(MacKenzie, 1993, p. 384)

Chapter 1: Introduction

Humans are social animals, and human emotions are fundamental to human communication, interaction, and cognition (Neufeld, 2015). As we step into an era where interaction with AI social robots and digital avatars becomes more common place, it is pertinent, even essential to "recognise the impact of actual events and the imprint of the social, historical, political, and relational matrix on the minds of [human beings]" (Cundy, 2015, p. xv-xvi). Indeed, the very foundations of our society may be in the process of being reshaped. The spiritual leader, His Holiness the 14th Dalai Lama (2012) has stated that "science and technology could replace the age-old spiritual and humanitarian values that have largely shaped world civilisation" (p. 4). The significance of this statement is considered in terms of AI social robots and/or digital avatars that will potentially fulfil a multitude of social roles in human society. This potential increases the necessity to understand the experiences people have had interacting with these AI social systems and how it will inform the future. Therefore, the purpose of the current study was to construct a theory that explains human beings' experiences and processes when interacting with an AI social robot and/ or digital avatar.

The methodology selected for the research was grounded theory; specifically, constructivist grounded theory as guided by Kathy Charmaz (2014). This approach to grounded theory has been widely used as a method and means of analysis in the investigation of social interactions and experiences (Birks & Mills, 2015; Charmaz, 2014). This qualitative, interpretive study investigated humans' experiences with AI social robots and/or digital avatars in an interactional context to determine what was happening and how it would inform the future. The outcome was achieved through the process of constructing a theory and applying the constructivist grounded theory-method package. It involved recruiting 15 participants who had personally interacted with an AI social system, and generating data through interviews, as well as theoretical sampling from extant data sources. The iterative process of constantly comparing and analysing data throughout the development of categories and subcategories led to the construction of a core theory for understanding human beings' experiences.

This introductory chapter presents the rationale for choosing the study, as well as the research aims and purpose to establish a roadmap for the rest of the thesis. Additionally, definitions and clarification of terms are outlined. This chapter also provides a

background context to position the research through my psychotherapeutic lens as the researcher and outlining my motivation for undertaking the study. The chapter concludes by summarising the structure of the thesis.

Definition of Terms

Herewith are definitions and clarifications of terms, used throughout the thesis, to give context and meaning to the discussion.

Human being is “a man, woman, or child of the species Homo sapiens, distinguished from other animals by superior mental development, power of articulate speech, and upright stance” (Encyclopedia.com, 2019).

Artificial Intelligence (AI) is “the branch of computer science that deals with the development of algorithms and techniques that can simulate or even recreate the capabilities of the human mind” (Vassilopoulos & Georgopoulos, 2010). Defined similarly in Oxford English Dictionary (2023) as, “the capacity of computers or other machines to exhibit or simulate intelligent behaviour; the field of study concerned with this”; AI is used in various types of technology, including automation, machine learning, deep learning, natural language processing (NLP), robotics, and self-driving cars (Techtarget, 2018-2019).

Social Robots “are robots that interact with humans and each other in a socially acceptable fashion, conveying intention in a human-perceptible way, and are empowered to resolve goals with fellow agents, be they human or robot” (Daily et al., 2017). Defined another way, a **social robot** “refers more precisely to artificial agents that are able to work in [various] fields by virtue of the ability to socially interact with human beings.” (Dumouchel & Damiano, 2017, p. 102). In other words, a social robot is seen as being a physically embodied entity that communicates and interacts with both humans and/or other agents in its environment by ensuing complex and dynamic social behaviours and rules connected to its purpose.

Human Robot Interaction (HRI):

> is the science of studying people’s behaviour and attitudes towards robots in relationship to the physical, technological and interactive features of the robots, with the goal to develop robots that facilitate the emergence of human-robot interactions that are at the same time efficient (according to original requirements of their envisaged area of use), but are also acceptable to people, and meet the social and emotional needs of their individual users as well as respecting human values. (Dautenhahn, 2013)

An avatar is “any embodied form for an AI [virtual] agent…embodiment in a human-like form [on a device] …because as human beings we like to connect with other human beings” (Frank, 2019). Defined another way an “AI avatar, also known as a **digital avatar,** are human-like bots that are created by AI-powered technology to increase human interaction. While digital avatars don’t only have a humanoid appearance, they can also communicate with people with the help of Natural Language Processing (NLP) algorithms” (Dilmegani, 2023).

Machine Learning is “a branch of AI trained on statistical models and algorithms, which enable it to make predictions and decisions. Using training and historical data, machine learning algorithms can improve and adapt over time, enriching its capabilities” (Zendesk, 2023). IBM (n.d.) defined machine learning as “a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy”.

Deep Learning is “a subfield of machine learning that structures algorithms in layers to create an ‘artificial neural network’ that can learn and make intelligent decisions on its own” (Zendesk, 2022). IBM (n.d.) defined deep learning as a:

> subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviour of the human brain—albeit far from matching its ability—allowing it to ‘learn’ from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy…Deep learning drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention.

Natural Language Processing (NLP), “in computer science, the use of operations, systems, and technologies that allow computers to process and respond to written and spoken language in a way that mirrors human ability. To do this, natural language processing (NLP) models must use computational linguistics, statistics, machine learning, and deep-learning models” (Ramanathan, 2023).

Artificial Narrow Intelligence (ANI) is :

> goal-oriented and designed to perform singular tasks and is very intelligent at completing the specific task it is programmed to do [i.e., facial recognition, speech recognition/voice assistants, driving a car, or searching the internet]. Artificial narrow intelligence (ANI), also referred to as weak AI or narrow AI, is the only type of artificial intelligence we have successfully realised to date…

> While these machines may seem intelligent, they operate under a narrow set of constraints and limitations, which is why this type is commonly referred to as weak AI. Narrow intelligence doesn't mimic or replicate human intelligence, it merely simulates human behaviour based on a narrow range of parameters and contexts (Engati, 2021).

Artificial General Intelligence (AGI), also called strong AI:

> is a computer's ability to do tasks commonly associated with human intelligence. The term is applied to the project of developing systems endowed with intellectual processes, such as the ability to reason, discover meaning, generalize, or learn from past experience. Computers can carry out very complex tasks with great proficiency (Britannica, T. Editors of Encyclopaedia., 2023).

Artificial Superintelligence (ASI) is "a software-based system with intellectual powers beyond those of humans across a comprehensive range of categories and fields of endeavour. ASI doesn't exist yet and is a hypothetical state of AI" (Barney, 2023).

Conversational AI "is a type of artificial intelligence that enables computers to understand, process and generate human language" (Cameron Hashemi-Pour, 2023).

> **Conversational AI** is a form of artificial intelligence that enables a dialogue between people and computers… This is achieved with large volumes of data, machine learning and natural language processing – all of which are used to imitate human communication … you can talk to your computer as if it were a real person. (Glover, 2023)

A **chatbot** is "a software or computer program that *simulates human conversation or 'chatter' through text or voice* interactions" (Brush, 2023). Microsoft (2023) defined a chatbot as "a software application that's used to engage in human conversation in a natural way. Chatbots are commonly used across many different industries for many different purposes".

To support flow of reading, in this thesis I have used the term **AI social system** interchangeably with the phrase 'AI social robots and/or digital avatars' for the purpose of reading flow and repetition. Additionally, AI social robots is also explained as '**AI robots**' and AI digital avatars is explained as '**AI avatars**' for the additional purpose of reading flow.

Rationale and Significance of the Study

As a practicing psychotherapist, I view people and how we relate with one another through a psychotherapeutic lens, of which some areas include attachment theory, trauma, and human drivers, as well as human suffering and need. Through conducting a previous study (Viljoen, 2017) into the impact of millennials' use of social media on their attachment experiences, I became more aware of the development of technology such as AI social robots and digital avatars. This technology has increasingly been promoted as having the purpose of 'helping humans be more humane' (CX Today, 2021) and address human needs that seem to be currently unmet, especially in light of the COVID-19 pandemic where there has been more physical and social isolation (The Conversation, 2021). When I began the study, popular media and researchers were focused on the study and marketing of AI robots and/or avatars across digital platforms. With the COVID-19 pandemic, the perceived need to attend to loneliness, mental health and other healthcare issues has been met with the solution of AI social systems (UneeQ, 2023). There has been research (Sciutti et al., 2018) regarding HRI and how to develop AI robots to 'fit in' with human beings' social ways and getting 'buy in' from people (including the academia that research this technology, the organisations that purchase the technology, the end-user, as well as general society and media). It was timely, interviewing participants before and during the COVID-19 pandemic as it provided insights into their perceptions of what need might be filled, or not, by this technology.

Neufeld and Mate (2013) stated that "technological innovations should be understood, not in terms of their content but in terms of how they change society. When we create a new technology, we are changing ourselves in fundamental ways" (p. 295). Consequently, this statement stimulated the question around what is happening now in our technological evolution with AI and a relationship with it, and how will this inform the future of human beings in terms of human relations, the role of this technology in human relations, and the effects on society at large. Additionally, in reflecting on the motivation for conducting this study, I considered Holmes' (2001) observation that, "our capacity to form relationships has been shaped by evolution and is mediated by the physical and biochemical architecture of the brain" (p. 21). This observation reflects the key concern around questioning what is happening in the area for human beings. This leads to the consideration of what is occurring and what is changing within human beings, when socially interacting with AI robots and/or avatars and the implications of how we work, play, and live. Consequently, interviewing people who have experienced interacting with AI social systems is fundamental rather than hypothesising about what their experiences might be.

This study began by asking how this type of social interaction - with an AI social system - affected the participants, what was happening for them, and how will it inform the future. With AI avatars being able to transcend different devices and platforms (including virtual reality and augmented reality, including holograms), it begs the questions 'What is reality?', how are human beings experiencing their interactions with AI social systems, and how does this influence the way humans live and connect?

As this research was conducted with a psychotherapeutic lens, Weizenbaum's (1976) (see initial literature review) deep concern around the misuse of AI as a therapeutic tool, in terms of ethical issues and possibility of devaluing human life, piqued my interest as both a researcher and a psychotherapist. Especially, in terms of the exponential development of AI social systems, specifically being promoted for social interactions with humans, such as in mental health, healthcare, and companionship (Soul Machines, 2019; World Health Organization [WHO], 2023). Cundy (2015) questioned whether people would turn away from face-to-face therapy that might be slow and potentially painful, and towards online quick-fix therapy (or AI social systems therapy). In considering Cundy's question, I reflected on authors such as Bowlby (1969) who stated that "there is [an] urgent need to be able to distinguish favourable development from unfavourable and also to know what conditions promote the one or the other" (p. 331). This statement seems as relevant now as when Bowlby wrote it; however, now the question could be, that in our contemporary times of the technological era what might the current 'conditions' be for relating and connection? Interviewing research participants and extant data to understand how this technological evolution and the resultant impact on human beings' processes and actions, as they experience it (both in terms of biological and psychological evolution), was critical in addressing the above questions.

Alenljung and Lindblom (2015) noted that while trends in HRI research acknowledge humans' user experiences interacting with AI robots and/or avatars as significant, it nonetheless fails to describe or further explore this topic. Alenljung and Lindblom suggested that current research paths in HRI, and specifically in human beings' experiences, have not been enough and further research is required. As such, the interest of the current study is from the subjective perspective of human beings themselves, rather than an attempt to investigate this from an external/objective perspective, and wondering about the exponential use of technology, regarding the human self, love, relating, play, thinking and experiences [of reality]. Therefore, the value and importance of interviewing participants who have experienced AI robots and/or avatars in an interactional context is central to study. The significance of the current study is to widen the lens of understanding and process on technology, specifically AI robots and avatars, as they will inevitably form a part of the everyday human society. This is an

area that is current and could potentially impact everyone, to some degree, personally and/or professionally, regarding technology, specifically AI social systems.

Human beings as a species, have evolved over thousands of years, and it seems our connection to each other has maintained our survival and prosperity (Neufeld, 2015). This connection and ways of relating can be both from quite primitive aspects of human physiology to more complex understandings, considering such things as language and culture/social context. While literature has begun to explore human beings' experiences when interacting with robots (Carpenter, 2016), it seems there remains a gap in the literature when considering, specifically, human beings experiences with AI robots and/or avatars. With these AI social systems continually developing mannerisms more closely resembling or mimicking human mannerisms; and, subsequently, being utilised as avenues to provide psychological therapy, medical, or other mental health support.

Research Aim and Purpose

The aim for this study was the construction of a substantive theory that identified and explained interactional experiences, from the human beings' perspective, with AI social systems—the process and actions, and subsequent impacts. The research aims to understand and conceptualise how human beings' experiences have been shaped, and the meanings they give to their experiences with robots and/or avatars, within the context of the 21st century (Aldiabat & Le Navenec, 2011).

The use of grounded theory methods, through a philosophical lens of symbolic interactionism and pragmatism, was a valuable method to construct new theory grounded from generated data. Additionally, acknowledging theoretical sensitivity recognised the details and the meanings of the data. To keep a wide net of sampling, to complement the grounded theory method of initial sampling (Charmaz, 2006, 2014), the inclusion criterion for recruiting participants were human beings who have had experience interacting with an AI robot and/or avatar. Subsequent theoretical sampling was guided by the data generated. As this study was underpinned by the philosophical perspectives of pragmatism and symbolic interactionism, it was dependent on the symbolic meaning people attribute to the processes of social interaction; thus, informing my understanding of how human beings' experiences have been shaped and the meanings they gave to their experiences with AI social systems.

The ontology was located in subjectivism and relativism, as it considered that views, aspirations, experiences, and behaviours were within the context of cultural assumptions, interests, and beliefs. This positioning focused on conceptualising descriptions and meanings of human beings' experiences interacting with AI social systems within their subjective context. The study was positioned in the epistemology of

social constructionism and constructivism as it furthered the assumption “that reality is constructed and reconstructed both individually from the sum of the experience and in relationship and conversation with others” (Ward et al., 2015, pp. 454-455), including the researcher as a co-constructor of both the research outcome and the research process. Thus, the epistemology was situated in a social context that considered knowledge and meaning as located within culture and history and contextually bound (Kuhn, 1970). As such, participants shared their experiences in interviews, and, as researcher and participant, we co-constructed the reality of their experience through the type of questions asked and in understanding meanings in their experience both at the time of the experience and, subsequently, during the interview.

The data gathering method incorporated interviews and extant data. Grounded theory methods consist of comparative analysis of data and developing codes and using memo-writing. The process of analysis and emergent theory explained processes and actions in relation to the research question, with the consideration that the theory generated in this research may be useful to other contexts. The scope of practice for this study were those people who have had experiences interacting with AI robots and/or avatars. These AI social systems have been gaining worldwide attention from academia, businesses, governments, and media (Broadbent, 2017; Carpenter, 2016; euRobotics AISBL, 2013; Soul Machines, 2019; The Conversation, 2021; WHO, 2022).

In summary, the aim of this study has been to construct a theory that describes the processes in relation to the research questions, with the view that the theory generated may be useful to other contexts. As such, it seems paramount that the findings of this study would be considered by AI researchers, technologists (who are ethically developing AI social systems in ways that account for human qualities that are especially involved in connection and relating), as well as the human end-users, the third parties that make use of these technologies to provide services for their customers, and for other professions such as psychotherapists and academia. We cannot afford to wait until the affects have happened to then consider them.

Study Context

After attending presentations given by Dr Neufeld, a developmental psychologist, one presentation stood out—“Raising children in a digital world” (Neufeld, 2015). This stimulated thinking and questioning around what is happening now in humans’ technological evolution and relationship with AI, and how it would inform the future for human beings, individually and socially. Neufeld and Mate (2013) stated that “technological innovations should be understood, not in terms of their content but in terms of how they change society.

When we create a new technology, we are changing ourselves in fundamental ways" (p. 295). These AI social systems have been increasingly promoted as having the purpose of 'helping humans be more human' and filling a gap due to lack of humans, or simply connecting data with rest of AI, or to society, or to help people. Therefore, their purpose has been promoted as addressing human needs that seem to be currently unmet, namely, in the COVID-19 pandemic, education, for the elderly, navigation, agriculture, stopping smoking, healthcare. The COVID-19 pandemic is a prime example where AI social robots have been designed to take people's temperatures or sanitise surfaces ongoingly; thus, human medical workers reduce their risk of being exposed to the COVID or putting their lives in danger (Shen, et al., 2020).

Motivation for the Research

The motivation for this study has been to widen the lens of human beings' understanding, processes, and experiences of AI robots and avatars in an interactional context, as these AI social systems will inevitably form a part of everyday human society in the not-too-distant future. Research has been done around HRI (Dautenhahn, 2007; Sciutti et al., 2018)—how these AI social systems need to 'fit in' with human beings' social ways and get 'buy in' from people (including from academic researchers, the organisations that purchase the technology, the human end-user, and media).

The COVID-19 pandemic has highlighted human needs and gaps in resources to service these needs; namely, for the elderly, addressing loneliness, healthcare, and mental health issues. This is a topic that is current and could potentially impact everyone, to some degree personally and/or professionally, with regards to these AI social systems. Results from a previous study (Viljoen, 2017) identified that online social media can become a barrier or a defence against engagement, particularly so when the boundaries between real and virtual are blurred when there is a loss of reality and transference displacement. Consequently, this understanding motivated the research question of how human beings' experience an AI robot and/or avatar, in an interactional context, to seek to understand these processes and experiences, and how it influences and shapes the way humans live and relate. Further, this study is a sociological endeavour because questioned the impact of how human beings relate as a society, and what this might mean for our future. A key point of concern in this research is, therefore, around the human experience of these AI social systems, how we are being affected, and the potential meanings of those affects. In reflecting on my motivation for conducting this research an observation from Turkle (2011) has been influential in my pursual of this topic; "a robotic face is an enabler. It encourages us to imagine that robots can put themselves in our place and that we can put ourselves in theirs" (p. 85).

This observation encourages consideration of the way human beings interact and socialise with others, and in particular the power of the face. Holmes (2001) too, makes a useful observation about the physiological and emotional, bases to human relating, that can change over time – moulded by evolution and facilitated by our physiology. Taken together, Turkle's suggestion that people are affected by a humanlike face, as well as Holmes' observation that human relating is influenced by evolution and is intrinsic to our existence, reflect my key motivation around questioning what are human beings' experiences and what is happening in this area.

Thesis Structure

This thesis follows the structure of a traditional PhD thesis, with the introductory chapter providing a background to the thesis topic and delineating the rationale and objectives for the study. **Chapter One** has introduced the research, defined key terms used in the thesis, and outlined the rationale and significance of the research, with consideration of the potential value of these findings. Furthermore, an overview of the aim and purpose of the research has been provided. Additionally, the context for the research has been delineated and my personal background, as the researcher, highlighted; including the motivation for the research.

Chapter Two presents the literature review from a constructivist grounded theory approach which includes reflecting on how initial literature was reviewed to establish a 'gap' in literature that could be explored. Then, a multi-phase literature review to subsequently identify literature to enlighten and expand the research is discussed. The literature reviewed initially included an outline of the development of AI technology over the years, the interactions of AI social systems with people, and concepts around the development of human beings from birth and the development of their social interactions and attachments.

Chapter Three outlines my methodological positioning and provides an overview of my paradigmatic position (qualitative and interpretive). I discuss the central ontological (subjectivism and interpretivism) foundations, followed by an outline of my theoretical perspective and the philosophical lens used in this study. I outline the key epistemological underpinnings (social constructionism and constructivism) that contribute to the choice of grounded theory methodology. I discuss the evolvement of grounded theory methodology and make specific reference to Charmaz (2003, 2006, 2014) version of grounded theory; namely constructivist grounded theory, explaining the applicability of this methodology for the research and topic.

Chapters Four and Five discuss the theoretical influences of Charmaz' (2006, 2014) theory-method package, illustrating the specific methods of gathering, generating, analysing and interpreting of data, and theorising, by way of research examples. Additionally, the challenges of interviewing participants on this topic and during the COVID-19 pandemic are discussed.

Subsequently, **Chapters Six**, **Seven**, **Eight**, **and Nine** present the research findings. **Chapter Six** presents the storyline of the constructed theory, '**Embodied Relating**', and provides a diagram to explicate the theory. **Chapters Seven** to **Nine** detail the three categories, subcategories, and the constructed grounded theory, as well as summarising the mechanisms that contribute to the outcome. **Chapter Seven** presents the first theoretical category, '**Unconscious Feeling: Thinking with feelings**', with the subcategories **Buying into the Hype**, **Having Expectations**, **Embodying**, **Developing Trust**, and **Growing Accommodating** outlined. **Chapter Eight** presents the second theoretical category, '**Conceptualising: Constructing of an Other**', with the subcategories **Projecting**, **Anthropomorphising**, **Getting in Flow State**, **Being Lulled**, and **Forgetting**. **Chapter Nine** presents the third theoretical category, '**Fantasy becoming Reality**', with the subcategories **Reality**, **Shifting Norms**, **Filling the Gap**, and **Fearing Misuse or Abuse**.

Chapter Ten presents a discussion of the theory, '**Embodied Relating**', and its meaning and value within the research area, as examined in the previous chapter. This chapter also includes extant theories that further substantiate and support the core theory of the thesis. Furthermore, this chapter summarises and reflects on the strengths and limitations of the study and proposes issues for further consideration for multidisciplinary fields and human end-users, such as the implications for further research in relation to the development and usage of these AI social systems by human beings, as well as the ethical and regulatory considerations.

Summary

This chapter has introduced the background context of the study, identifying the two key technologies, namely AI robots and avatars, and how human beings' experience interacting with them. I have outlined my aim and purpose for constructing a substantive grounded theory that identifies and explains the processes of interaction and relating that occur between human beings' and AI social systems, and the value in learning more about the human experience of interacting with these technologies. So, at the very least, to 'do no harm'; and, at the most, provide an understanding on how best to utilise these AI social systems (from the researcher to technologist to the third party purchasing the technology to the human end-user) by providing the intended support or regulations that

humans require. This introduction has delineated the necessity for exploring human beings' experiences interacting with AI robots and/or avatars. Additionally, the selected methodology for the study has been outlined. The following chapter identifies relevant initial literature, providing a foundational context for the research, and explains how this research fills a gap from the 'human-experience' perspective. Whatever thoughts or views one has of technology, it seems critical and timely to endeavour to discover what is happening and to be willing to confront these technology's effects on us.

> We shall have to come to terms right now with the fact that the boundary between trickery and honest tries is imprecise; this fact too is part of the history of thinking machines…From the start, the tangle between fact and fiction has been difficult to unravel. (McCorduck, 2004, p. 7)

Chapter 2: Literature Review

Introduction

In this chapter, I discuss the literature review approach undertaken, highlighting how the literature informed and, later, substantiated the results derived from the data generation and analysis. I begin with a discussion on the place of literature in grounded theory research and outline the method of the literature review used in this study; namely, a multi-phase literature review (Thornberg & Dunne, 2019). The initial phase of the literature review (discussed in this chapter) was to determine whether the research topic had been previously studied, and reasons for that; thus, allowing me, as the researcher, to familiarise myself with the topic more broadly. Specifically, I outline a scoping review of the development of AI and subsequent fields, such as what has come to inform what is now known as HRI and other related AI social technology (i.e., digital avatars) relevant to the research question. Additionally, as this study is about human interactional experiences, I reviewed extant literature relevant to human development, connections, and socialisation. The chapter concludes with an outline and explanation for the value of the study in addressing a gap in the literature.

The ongoing and final literature searches will be outlined in the Discussion, Chapter 10. The ongoing review resulted from engaging with literature that became relevant during data generation and analysis, and that I had not anticipated at the outset of this study. Additionally, in contemporary grounded theory, literature can be used to further enrich data analysis, be considered as a data source for theoretical sampling, and a tool to stimulate theory construction (Birks & Mills, 2015; Charmaz, 2014; Thornberg & Dunne, 2019). The final literature review focuses on extant literature which was guided by categories and concepts in the theory developed in this study; thus, locating, evaluating, and defending the theory by positioning the research findings in the context of current literature (Charmaz, 2014; Holliday, 2002).

The Position of Literature in Grounded Theory Research

Grounded theory literature review has proven to be a contentious issue (Dunne, 2011; Thistoll et al., 2016; Thornberg, 2012; Thornberg & Dunne, 2019) which fundamentally ensues from Glaser and Strauss' (1967) original and controversial phrase '*tabula rasa*' or 'blank slate'. Glaser's (1992) perspective put forward that the researcher should

delay the literature review (until data analysis was completed) and focus exclusively on the area of study in order to be uncontaminated by existing theories or knowledge (Birks & Mills, 2015; Charmaz, 2014, 2006). In other words, recommending against conducting an early literature review out of concern that the researcher may be unduly influenced by existing theoretical frameworks and either consciously or unconsciously applying these onto the grounded theory research process (Dunne & Gamze, 2020). While collaborating with Corbin in the 1990s, Strauss adjusted his stance; consequently, Strauss and Corbin (1990, 1998; Corbin & Strauss, 2015) acknowledged that researchers bring with them a combination of prior experiential and discipline-relevant understanding of their field and existing studies - thus, enhancing the researcher's sensitivity to the data being studied (Bryant, 2009; Dunne, 2011; Strauss, 1987; Strauss & Corbin, 1990, 1998). However, Strauss and Corbin (1990) also highlighted that it is not necessary to complete a full review of literature initially, in order to allow the data to emerge and not limit theory development (Thornberg & Dunne, 2019).

As grounded theory methodology has developed, the original classic grounded theorist's perspective of delaying the literature review until after data analysis to avoid contamination of the research process has been argued by numerous contemporary grounded theorists as insufficient, flawed, and unrealistic (Charmaz, 2014; Clarke, 2005; Dey, 1999; Dunne, 2011; Goldkuhl & Cronholm, 2010; Thornberg, 2012). Dunne and Gamze (2020) questioned how a researcher would be certain of the "boundaries and parameters of our own substantive area without first familiarising ourselves with it" especially if the nature of the research is interdisciplinary. Dunne and Gamze highlighted that in order to determine if there is a gap in the literature on a topic that warrants investigation (Creswell, 1998), the researcher has to engage with existing literature, stating "what other way is there to conclude that there is indeed a lack of information on a given phenomenon without firstly undertaking an initial literature review?" (Dunne & Gamze, 2020). Additionally, Dey (1993) contended that "to analyse data, we need to use accumulated knowledge, not dispense with it. The issue is not whether to use existing knowledge, but how" (p. 63). Furthermore, Dunne and Gamze considered how not reviewing literature in the field to be researched, before data collection, might result in insufficiently informed choices regarding the research aims, objectives, and research questions.

Charmaz (2014) contended, in line with constructivist philosophy, that research does not occur in a vacuum; instead, is shaped and informed by the researcher's perspective in which they are operating (Kenny & Fourie, 2015).

Consistent with the constructivist and pragmatist viewpoint, Charmaz (2008, 2014) acknowledged the issue is not to ignore literature and existing knowledge; rather, identify these positions and evaluate their effect on the research process by subjecting them to rigorous examination. In so doing, the researcher can work consciously at developing theoretical sensitivity during the research process, thus valuing the researcher's "tacit knowledge and the existing literature as 'sensitising concepts' within the research process" (Ward et al., 2015, p. 457). In acknowledging Herbert Blumer's (1969) notion of sensitising concepts, grounded theorists begin their research guided by specific research interests, sensitising concepts, and disciplinary perspectives. These sensitising concepts are used as tools to provide initial, yet tentative, ideas (as a starting point) for researchers to pursue and sensitise to certain questions on the topic under study (Bryant, 2017; Charmaz, 2014). Charmaz (2014) stated that "a sensitizing concept is a broad term without definitive characteristics; it sparks your thinking about a topic" (van den Hoonard, 1997, p. 30). Accordingly, sensitising concepts form part of the studying of research data, coding throughout successive levels of analysis and theoretical category development from the data (Bowen, 2006, Charmaz, 2014).

The initial literature review was valuable in establishing a valid rationale and determining sensitising concepts (e.g., AI, HRI, and attachment theory). These sensitising concepts were guided by my tacit knowledge, prior experiential and discipline-relevant knowledge within the psychotherapy field, as well as concepts collected from reviewing existing literature; informing my starting position and acknowledging my 'psychotherapeutic lens' of the study by engaging with the relevant literature throughout the research process. Thus, sensitising me to emerging ideas and contributing to theory construction. Kelle (1995, 2005) contended:

> that pre-existing theories and research findings can be used as heuristic tools, i.e., using extant concepts, theories and ideas as "lenses" and tools that help the researcher to focus the attention on certain phenomena, aspects or nuances as well as imaginarily see beyond data. (Thornberg, 2012, p. 8)

In this study, extant theoretical concepts and ideas from literature, within the wider field, have to make sense to fit with data and substantive codes and concepts. Therefore, by remaining aware of sensitising concepts and examining the influence of the extant theoretical codes and concepts, they needed to earn their way into this grounded theory research (Thornberg, 2012). By acknowledging my prior knowledge, theoretical influences, and analytic lenses, I recognised the need for constant reflexivity in order to remain grounded in the data (Mills et al., 2006a).

To aid this process of reflexivity and self-monitoring, I used memoing as a tool (Dunne, 2011) to reflect on and become aware of how I constructed my own concepts and compare these with the multiplicity of theories and extant concepts in the literature (Mills et al., 2006a; Thornberg, 2012). Furthermore, as part of my reflexivity I completed a pre-supposition interview at the beginning of the research process which highlighted my preconceived ideas. I then used discussions with my researcher supervisors throughout the research process (Birks & Mills, 2015; Charmaz, 2014; Thornberg, 2012). These reflexive strategies are discussed further in the Methods, Chapters 4 and 5.

The constructivist grounded theory approach offers flexibility in regard to early literature engagement (Dunne & Gamze, 2020). Thornberg (2012) noted that using literature deepens the analysis. The researcher can concurrently take the viewpoint of critiquing and challenging "emergent" (Thornberg, 2012, p. 4) concepts and thinking; as well as use constant comparative analysis of both data and literature to expand, refine, or critique pre-existing knowledge and extant theories. Thornberg referenced informed grounded theory as "a product of research as well as the research process itself in which both the process and the product have been thoroughly grounded in data by GT [grounded theory] methods while being informed by existing research literature and theoretical frameworks" (p. 249). Thornberg recognised its origins in constructivist grounded theory and the pragmatist concept of abduction, and acknowledged the researcher's position within a historical, ideological and socio-cultural context wherein the data are always socially constructed and not an accurate representation of reality. This study is guided by Charmaz' (2006, 2008, 2014) constructivist grounded theory and incorporates Thornberg's concept of informed grounded theory.

While Charmaz (2014) recommended a designated literature review chapter, she also promoted that literature be incorporated and interwoven into all phases of the grounded theory research. Employing literature throughout the research and balancing it with delaying in-depth literature immersion until after the core categories had been constructed, allowed for creativity and discovery—essential components of grounded theory (Kenny & Fourie, 2015). Further, in considering what literature is appropriate to engage with, at what point in the research process, Thornberg and Dunne (2019) suggested that as the researcher engages with extant literature throughout the research process, there are particular stages where distinct literature is sought, explored, and incorporated in the research.

Multi-phase Literature Review

Constructivist versions of grounded theory emphasise the importance of the researcher maintaining constant engagement with the extant literature throughout the whole research process (Thornberg & Dunne, 2019), which complements the iterative strategies when moving back and forth between data and analysis (Charmaz, 2014). This process is important with grounded theory studies and is reflective of the constructivist grounded theory methodology. Therefore, intertwining the three different phases of the literature review allows for specific types of literature to be sought, explored, and integrated into the research; thus, remaining involved with the data and emerging analysis. This process then contributes to the research by enhancing theoretical sensitivity when analysing data and developing theoretical codes. (Birks & Mill, 2015).

This study applied the multi-phase literature review in grounded theory, as identified by Thornberg and Dunne (2019) which are: initial, ongoing, and final literature review. These are each distinct and purposeful phases, serving valuable functions in the research process (Thornberg & Dunne, 2019). As previously mentioned, my initial literature review is outlined in the following sections of the current chapter. The two consecutive literature review phases; namely, ongoing literature review (carried out during data analysis) and final literature review (being guided by categories and concepts in the theory and to defend the theory) have been primarily discussed in the Discussion, Chapter 10.

The first phase of the literature review commenced at the early stage of the research process, prior to data collection and data generation, and helped establish the “if, how, when, where, why and by whom a given topic has been studied (or not) to date” (Thornberg & Dunne, 2019, p. 211). Thus, I familiarised myself with extant literature, identifying gaps that justify exploration and locating myself in this field of study. Furthermore, the initial literature review informed my data generation (i.e., participant interview questions) and guiding analysis. As a result of completing the initial literature review, I was able to provide a sound rationale for the study; contributing an original piece of research and meeting the university research proposal requirements (Birks & Mills, 2015; Bryant, 2017; Charmaz, 2014; Urquhart, 2007). In the initial literature review I did a scoping of literature in general to gain a widespread perspective of the development of AI and subsequent fields, such as what has come to inform what is now known as HRI and other related AI social technology (i.e., digital avatars) relevant to the research question.

Thus, the initial literature review gave me a framework for understanding the types and development of AI social technology by reviewing relevant texts to learn about key drivers and turning points in technology over time, and more recently. Furthermore, I searched for literature on HRI which emphasised a gap in literature in terms of exploring human beings' experiences when interacting with social robots, with an even more substantial gap in literature around Human-Digital Avatar Interactions (HDAI) and human beings' experience interacting with AI digital avatars (discussed at the end of this chapter). Additionally, as this study is about human interactional experiences, I reviewed extant literature relevant to human development, connections, and socialisation. This initial literature review resulted in generating sensitising concepts as I moved through the research process, facilitating my creativity (Charmaz, 2014).

In the second phase of the ongoing literature review, I continued to engage with the literature while undertaking the primary data collection (i.e., collecting raw data from qualitative interviews), which directed me to what type of literature to draw on (Thornberg & Dunne, 2019). The literature, strongly informed by my data, emerging codes and categories, and the constant comparative process, allowed me to compare emerging concepts from data with existing studies and theoretical frameworks in order to expand the emerging theory. The iterative process of moving between data and literature was guided by the technique of theoretical sampling (Thornberg, 2012), as suggested by a several authors who support the use of theoretical sampling when the data directs the researcher to do so (Birks & Mills, 2015; Charmaz, 2014; Thornberg & Dunne, 2019). Seeking out existing theoretical constructs enriched the iterative analytical process, enhancing sensitivity to subtle nuances in data and contributing to the development of my grounded theory (Dunne & Gamze, 2020).

As the ongoing literature review moved into the final literature review phase and data collection, data generation, and analysis progressed; I sought to identify existing theories and concepts to help elevate the grounded theory to a more abstract level in theory construction, with ideas or concepts either earning their place or being discarded (Birks & Mills, 2015; Charmaz, 2014; Thornberg, 2012). An example was that I had not understood, when starting out with the literature review and research process, how much embodiment formed part of a human beings' interactional experience. Through analysis of data collected from interviewing participants who had an interactional experience with a robot and/or avatar, it became progressively evident how and when humans used their bodies to interact and respond with the AI social system. This motivated me to search further literature to understand and explain what I was seeing in the data, ascertaining when emergent concepts were useful my theory construction, and discarding those that were not (Thornberg, 2012).

The purpose of the final literature review was to support refinement of contribution to knowledge to the field (Thornberg & Dunne, 2019) and assist in more formally contextualising and situating the study within and across disciplines (Charmaz, 2006) during the later stages of analysis, theoretical coding, and theory development (discussed in Methods, Chapters 4 and 5; and Discussion, Chapter 10) (Birks & Mills, 2015; Bryant, 2017, 2021; Charmaz, 2014; Glaser & Strauss, 1967; Thornberg & Dunne, 2019). I found I was drawn towards theoretical ideas which historically have not been applied in the given disciplines discussed in the current research, but which resonated powerfully with the data (Dunne & Gamze, 2020).

In sum, conducting a grounded theory constructivist approach, in conjunction with applying a multi-phase literature review, allowed me to harness the advantages of conducting an initial, ongoing, and final literature review without 'forcing' it on the data (Thornberg & Dunne, 2019). Each of these three phases were targeted and purposeful; yet individually they fulfilled specific and valuable functions in the overall research process (Thornberg & Dunne, 2019)

Initial Literature Review

In order to have a sense of how these AI social systems came to be, it was important and worthwhile to begin a literature review identifying the human interest in developing artificial creatures and the technological developments that followed. Furthermore, as this study is about human interactional experiences, I reviewed extant literature relevant to human development, connections, and socialisation to bring these two fields into dialogue. This initial literature review also considers the most salient studies and gaps in literature and provides a rationale for this research.

The Development of Artificial Creatures

In ancient times (850BC) there were myths and tales of artificial creatures gifted with intelligence or consciousness, and so AI began with an ancient wish to copy the Gods in acts of creation (McCorduck, 2004). Capek (1921) used the Czechoslovakian term 'robot' (meaning drudgery, servitude, and referencing the idea of forced labour) in his science fiction play '*R.U.R.*' (in English meaning 'Rossum's Universal Robots'), which introduced the word 'robot' to the English language and to science-fiction as a whole (Asimov, 1979). Over the years, there have been ideas about what a robot is, how it is defined, and the meaning of the word robot. For example, if a person is referred to as 'a robot', it tends to be as a result of the way the person acts and responds, namely, "*in a mechanical, routine manner*" (Dictionary.com, 2023).

Isaac Asimov, a well-known science fiction author, wrote in some of his stories about robots, such as a robot who wanted to be a man; another about a female robot; and

another about a robot buying his freedom; as well as Norby, the mixed-up robot for children (Asimov, 1995). Science fiction anthologies appeared by the hundreds and many of Asimov's stories were among them. Asimov (1995) was of the opinion that science fiction began in the late 19th century with popular French writer Jules Verne, whose writings included '*From Earth to the Moon*' (1865) and '*Around the World in Eighty Days*' (1873). Verne was followed by other science fiction who were less well-known but later became popular, such as Herbert George Wells ('*The Time Machine*' (1895) and '*The War of the Worlds*' (1898)); followed by other science fiction books, such as Aldous Huxley's (1932) '*Brave New World*' and George Orwell's (1984) '*1984*'. Generally, science fiction readers of the 1930s and 1940s tended to consider only the magazines and ignored the occasional literary novel. The science fiction stories in magazine form were, for the first 12 years, action-oriented and written largely by authors who knew little or nothing about science. During World War II, science fiction abruptly changed. After 1938, characters being real scientists and engineers, and conversed naturally on these topics, with science fiction stories becoming more idea- and puzzle-oriented (Asimov, 1995). Science fiction stories included nuclear bombs, German rockets, and the hopes of space flight and electronic computers, which were the staples of science and became reality in the post-war period (Asimov, 1995).

The beginnings of what is now the 21st century AI, was established by classical philosophers and early thinkers (McCorduck, 2004) who endeavoured to elucidate the formula of human intelligence as the mechanical manipulation of symbols. The invention of the programmable digital computer in the 1940s was the result of this early work and inspired a small group of scientists to earnestly begin contemplating the option to create an electronic brain. Warren McCulloch and Walter Pitts (1943) are largely acknowledged as having done the first work in AI, drawing on three sources: "knowledge of the basic physiology and function of neurons in the brain; a formal analysis of propositional logic due to Russell and Whitehead; and Turing's theory of computation" (Russell & Norvig, 2021, p. 35). Alan Turing (1950), a British mathematician and logician, argued that a "thinking machine" (Norvig & Russell, 2003, p. 948) was possible and he developed the 'Imitation Game' in 1950, which later became known as the Turing Test. Using a simple, rational method to evaluate whether machines can think, with the assumption that it would not be obvious for a person to tell if they were communicating over a teleprinter, with a human or a machine, in 1955 McCarthy and colleagues coined the term 'AI' and the Turing Test subsequently became the first significant assertion in the philosophy of AI. Already in 1965 there were concerns about the idea that machines could become 'ultra-intelligent', and Professor I.J. Good (year) coined the term an 'intelligence

explosion', a theoretical consequence in which AI would become so powerful that it would pose a danger to humanity.

Joseph Weizenbaum (1976) developed a natural processing computer program between 1964 and 1966 named 'ELIZA', a time-shared computer system (systems in which multiple users interact with a single mainframe computer) capable of language and interaction that could be 'taught' to speak increasingly well. Weizenbaum collaborated with Kenneth Colby (a psychiatrist) to design the 'DOCTOR' programme which answered simple questions (McCorduck, 2004). However, Colby et al. (1966) created a computer programme, based on ELIZA, with the ability to conduct psychotherapeutic dialogue. Weizenbaum was deeply concerned about the ethical principles of this programme and was reportedly "disturbed that Colby saw a mindless program as a serious therapeutic tool" (p. 5). Weizenbaum's differing views of good science and ethical behaviour resulted in Weizenbaum publishing the book '*Computer Power and Human Reason*' which argued that the misuse of AI has the potential to devalue human life. As this research was conducted with a psychotherapeutic lens, Weizenbaum's deep concern piqued my interest as both a human researcher and a psychotherapist. It contributed to my focus on the topic, with the exponential development of AI social systems, specifically being designed for social interactions with humans, such as in mental health, healthcare, and companionship (by companies like Hanson Robotics, 2023; Softbank Robotics, n.d.; Intuition Robotics, n.d.; Soul Machines, 2023; Replika, 2023; Being AI, 2021).

Marvin Minsky, an American mathematician and computer scientist, was inspired by work on computer vision, generalised to language and other tasks requiring intelligence (McCorduck, 2004). Minsky defined AI as "the science of making machines do things that would require intelligence if done by men" (Stonier, 1992, p. 107).

Licklider, a psychologist who was central in the early development of AI, wrote a series of memos outlining his first concepts in 1962, of what eventually developed into the Internet, "a worldwide medium for collaboration, information dissemination, broadcasting, and interaction between individuals, regardless of geographic location" (McCorduck, 2004, p. xxvii). In 1967, AI underwent a scientific milestone in determining that knowledge was considered to be as valuable as reasoning in intelligent behaviour (McCorduck, 2004). The commercialisation of AI started in 1981, with the publicised project (Fifth Generation) which had important AI goals, such as "writing programs and building machines that could converse, translate languages, interpret pictures, and reason like human beings" (Crevier, 1993, pp. 211-212), funded by the Japanese Ministry of International Trade and Industry (with other countries subsequently developing their own programmes).

Significant achievements included, in 1985, R. Brooks' demonstration of 'Allen', the first of his autonomous reactive robots; and then in 1987, Minsky wrote a book '*Society of Mind*' (portraying the mind as a 'society' of tiny components that are themselves mindless); and Berners-Lee begun working on the World Wide Web at Cern, in Geneva in 1988 (McCorduck, 2004). Following these earlier developments and achievements in technology, and premature industrialisation of AI, commercial collapse subsequently resulted. The optimistic and highly promised expectations declared by AI pioneers appeared unreachable, failed to be realised, and financial funding disappeared. The challenging years that followed, during most of the 1980s, were known as the 'AI winter' (McCorduck, 2004). Subsequently new ideas were explored, leading to AI being beneficial to solving many challenging problems (Crevier, 1993). By the late 1980s, several researchers advocated a completely new approach to AI based on robotics (McCorduck, 2004). They believed that to show real intelligence a machine needed to have a body with sensorimotor skills and advanced abilities in common-sense reasoning, which includes perception and movement, to successfully navigate within the world. Moravec (1988) stated that "fully intelligent machines will result when the metaphorical golden spike is driven uniting the two efforts [real world competence and common-sense logic]" (p. 20).

AI Today

The area of AI, more than half a century old, has been successful in eventually meeting some of its original goals that had captured the imagination of the world since the 1960s, such as the vision of human level intelligence in AI. Investment and interest in AI surged in the 21st century, with machine learning being successfully applied to numerous challenges in academia and industry, as a result of availability of powerful computer hardware (AI New Zealand Ltd, 2023). An example was a significant event for AI, where, in 1997, the AI technology 'Deep Blue' defeated the chess world champion Garry Kasparov; followed by further AI developments with the commercialisation of robot pets and smart toys in 2000 (i.e., Kismet – a robot that could display emotions) (McCorduck, 2004). Success in the AI field was partly due to increased computer power and focusing on particular problems; consequently, split into competing subfields focusing on specific challenges or approaches (Russell & Norvig, 2021). While AI has not always been visible to the general public, it has been providing solutions. The development of machine learning has been applied to a number of challenging problems such as, data mining, speech recognition, and medical diagnosis (Olsen, 2006).

AI is a division of computer science that attempts to understand the essence of intelligence and build new intelligent entities that respond in a manner similar to human intelligence.

Research in this area includes robotics, speech recognition, image recognition, NLP, and expert systems (mimicking the decision-making ability of a human expert) (Kurzweil, 2005; Russell & Norvig, 2003). Since the inception of AI, the theory and technology have become more and more developed, with the application fields having increased and the consequent ability of AI to imitate the information process of human consciousness and thinking.

Machine learning is an area of study considered to be a part of AI, and is dedicated to using machines to understand and building methods that 'learn'; in other words, processes that leverage data to enhance performance on a certain set of functions (Mitchell, 1997). Deep learning is a branch of machine learning. It is a model developed to continuously analyse data with a logical structure, thereby achieving intelligent decisions on its own, using many processing layers of algorithms called an artificial neural network. The design of an artificial neural network is guided by the biological network of neurons in the human brain, resulting in a learning system with increased capabilities (LeCun et al., 2015; Zendesk, 2022). While deep learning and machine learning operate in similar ways, their proficiencies are different, with machine learning models increasingly better at performing specific functions as it acquires new data. However, it still requires some human intervention. A deep learning model uses algorithms to determine the accuracy of calculations through its own neural network, without requiring human intervention. As a result, it is what powers the most human-like AI (Zendesk, 2022). Another area of AI is language processing engines that can competently defeat humans at answering general trivia questions, such as the AI called IBM Watson with Deep Blue.

Where the use of AI has developed is from more basic algorithms to assist background programming functions, gaming, and more applied AI, and has grown out of a business focus to capture consumers for marketing, finance areas, gaming purposes and, more recently, various aspects of the health sector. However, through machine learning, AI has been increasingly used to interact in with humans, referred to as 'social machines'. As such, there is popular writing and discussions around the use of AI robots and avatars, in more social settings with people. Developers of these types of AI (e.g., Hanson Robotics, 2023; Softbank Robotics, n.d.; Intuition Robotics, n.d.; Soul Machines, 2023; Replika, 2023; Being AI, 2021) have marketed them as assisting and benefiting humans in various social settings; suggesting, perhaps, providing a better outcome than another human would. David Hanson, from Hanson Robotics (SDG, 2022), suggested that as the AI develop empathy and care for sentient beings (presumably humans) so they will be able to fulfil caring roles, such as caring for elderly, children, and providing psychological therapy for people.

A digital avatar company follows a similar aim, stating that their digital avatars include these attributes “privacy; security; ethical AI; ability to provide a multi-modal experience; expression of the digital person’s emotions while also observing and empathetically reacting to users’ emotions; and manifesting a personality” (Soul Machines, n.d.). Findings from Broekens and Heendrik’s (2009) study indicated “a positive effect of companion type robots in health care for the elderly with respect to at least mood, loneliness and social connections with others” (p. 100). Fiske et al. (2019) conducted research into embodied AI having an increased clinical relevance for therapeutic applications in mental health services; that is, in psychiatry, psychology, and psychotherapy. These include technology from ‘virtual psychotherapists’ to social robots in dementia care and autism disorder and robots for sexual disorders; as well as identifying the aim of embodied AI application in mental healthcare, to improve the quality of care and to control expenditure. It is interesting to note that through this initial scoping review, I found limited research and discussion papers around the area of AI and attachment (Carpenter, 2013; Coulson et al., 2018; Law et al., 2022; Rabb et al., 2022; You & Robert, 2017); however, I did not find research conducted on the topic, as discussed in the current study.

Within the field of AI, there are subfields of applied and generalised AI. Applied AI is the most widespread and is developed to action distinct functions based on data and rules, such as moving of a self-governing vehicle. In theory, generalised AI (not as widespread yet) can handle any function and has led to the development of the field of machine learning; established on the theory that if the AI is provided big data, it can calculate and solve a broad range of tasks on its own (Digital.govt.nz, n.d.). Some of the influential founders of AI (Beal & Winston, 2009; McCarthy, 2007; Minsky, 2007), suggested that the aims of AI should focus on “machines that think, that learn and that create” and “a machine should be able to learn to do anything a human can do” (Russell & Norvig, 2021, p. 50).

Another endeavour with similar goals, was that of the AGI initiative, as discussed by Goertzel and Pennachin (2007), referring to machines than would have the capability of experiencing consciousness. At a similar time, apprehensions were being expressed about the development of artificial superintelligence (ASI), which is when machine intelligence far surpasses human intelligence and abilities, being a misjudgement (Omohundro, 2008; Yudkowsky, 2008). Turing (1996) expressed similar concerns and referenced the science fiction novel ‘*Erewhon*’ (Butler, 1872). Butler’s novel was inspired by his years living in colonial New Zealand and having read Darwin’s ‘Origin of Species’; a satire on conventional virtues, religious hypocrisy, and the unthinking acceptance of beliefs.

These concerns have developed extensively since more recent developments in deep learning with public statements being made by Stephen Hawking, Bill Gates, Martin Rees, and Elon Musk (Bostrom, 2014; Russell & Norvig, 2021).

AI technology is becoming an increasingly prevalent technological tool in the 21st century, where the use of AI seems to have developed from more basic algorithms to assist background programming functions such as in calculations or summaries of data, in the gaming industry, and in more applied subfields of AI (McCorduck, 2004). AI development has been motivated by certain problems or approaches, such as being utilised as a customer service tool in various platforms. By 2016 (Lohr, 2016), there was significant spikes in interest for AI-related products which, subsequently, drove advances and progressed deep machine learning research, such as robotics, speech recognition, image recognition, and NLP (LeCun et al., 2015). However, AI has also been increasingly developed to interact with humans (Goertzel, 2018; Sciutti et al., 2018), referred to as social machines, in social settings where advances in AI and robotics are becoming progressively more useful in everyday lives Developers of these types of AI, such as at Hanson Robotics, have marketed these AI as assisting and benefiting humans in various social settings (Goertzel, 2018). As a part of my scoping review, I attended a conference in Barcelona, Spain (5th HR Conference, 2018) around the topic of AI and human interaction, a current initiative which included people engaging with 'Sophia', a AI social robot developed by Hanson Robotics.

Sophia, herself, explained that an advantage of not having 'human emotions' is that 'robots do not judge' (5th HR Conference, 2018). A belief reflected in Sophia's creator, David Hanson, who has been quoted as saying that he "believes that three distinctively human traits must be integrated into the artificial intelligence of these social robots: creativity, empathy and compassion" (Radii, 2018).

One technology company is developing AI digital avatars and referencing these AI social systems as **"**emotionally intelligent artificial humans" and "incredibly life-like" (Soul Machines, 2019) with their own form of a central nervous system that imitates the capabilities of the human brain to make it more realistic (Strang, 2017). **T**hese AI avatars reportedly have the capacity to be emotionally responsive and intelligent, even imitating empathy. Some digital avatar companies have claimed that their **"**vision is to humanize artificial intelligence to better humanity" (Soul Machines, 2019) through the use of digital avatars and could be utilised in customer services roles, mental health/therapeutic roles, in education at schools, and as digital healthcare assistants. Another has claimed that their AI social system allows "users [to] express themselves in a safe and nurturing way, 'allowing you to engage with your most emotionally connected self' (Prendergast, 2019); and in popular media a claim that an AI digital avatar has been "designed to meet [human

users] needs… help[ing] them to share their emotional problems, thoughts, feelings and experiences" (Schön, 2021). Yet another digital avatar company claimed their digital avatars "offer friendship, entertainment, learning experiences and adventure… [and] bring humanness to humans… to help us thrive" (BeingAI, 2021). Fiske et al. (2019) highlighted embodied AI as an optimistic perspective across the field of mental health, while acknowledging a marked need for more substantial research, including investigating the benefits and potential harms of current and future AI applications, especially in regard to relational contexts between human beings and AI. This raises the question, what actually are the experiences of human beings who are interacting with these technologies, and what does this mean for the future?

HRI

HRI is a topic that has multidisciplinary contributions from fields ranging from social sciences, robotics, AI, as well as natural language understanding and design (Dautenhahn, 2007). The concept of HRI started out many years before as science fiction, with authors writing novels about robot and human interactions (Asimov, 1950). Research, identified in this initial literature review, has been conducted around how to make the robots become more capable and autonomous, perhaps more 'human-like' in thinking and actions, so that humans may be more motivated to interact with robots (Bartneck et al., 2020; Goodrich and Schultz, 2007; Kanda & Ishiguro, 2013; Parasuraman et al., 2011). Other literature claims to take a different perspective, suggesting that it is not the human that needs to adapt to the robots, rather that robots need to:

> …adjust to their human interaction partners in better ways. What do the robots have to learn to be considerate of people and, no less important, be perceived as considerate by people? Which skills do they need, what do they have to learn to make cooperation with humans possible and comfortable? (Sciutti et al., 2018, p. 23).

The literature stated that the aim of their research on HRI was to humanise interactions, which "implies a mutual understanding between the two agents, with the effort of adapting to the partner not falling on the human shoulders alone, but rather being shared between the two" (Sciutti et al., 2018, p. 22). It seems, from this research area, there has not been an exploration as to the impacts and experiences of these HRI on humans; neither the motivations nor reason behind humans' expectation for HRIs. This seems like a significant and important gap in the research and literature. Carpenter (2016) conducted research and wrote literature around emotional attachment to robots in military settings, which produced findings that soldiers can develop emotional bonds to robots deployed in place of humans.

Turkle (2011) also referred to how quickly and deeply people of a variety of ages seem to attach to robots of various shapes, such as 'Kismet' and 'Cog'. While these studies and literature are exciting, and beginning to consider humans' experiences when interacting with robots, there remains a gap in the literature when considering human beings' experiences with AI robots and/ or avatars. Alenljung and Lindblom (2015) noted that while trends in HRI research acknowledge human-user experiences interacting with AI social systems as significant, it nonetheless fails to describe or further explore this topic. Alenljung and Lindblom suggested that current research paths in HRI, and specifically in human beings' experiences, have not been enough and further research is required. In this initial literature review, I explored existing knowledge to gain understanding of human beings, in terms of socialisation, psychology, and relating, and began by noting McCorduck's (2004) (an author immersed in the milieu of the great and greatest in AI at Carnegie Mellon) statement:

> history of psychology is nothing less than a history of the effort to explain what many people still regard as ultimately inexplicable, the workings of the human mind… the notion of mind slowly gave way to the notion of brain function… if you consider thought to be a mechanical process, you can view human beings as machines, though of a very special kind. (p. 37)

This quote by McCorduck (2004) reflects a perspective of considering the human mind and thought as a mechanical process; thus, viewing humans as machines, like a metaphor for a computer. It forgets or ignores other aspects of being human—of relating, feeling or being. In this initial literature review I now consider human beings from a psychotherapeutic lens: how we as humans relate and connect with each other, as a starting point for exploring how humans experience interacting with AI robots and/ or avatars. Through research we can begin to gain clarity and understanding of human beings' experiences are and what is happening, when interacting with AI social systems. Harari (2018) posited that "humans have bodies. During the last century technology has been distancing us from our bodies" (p. 88) and that "what people might really need are the tools to connect to their own experiences" (p. 89). These statements resonate with my findings that human beings are being affected by technology advances more recently, in fundamental ways.

Looking at Human Beings and Our Interactions

To understand human beings' experiences when interacting with artificially intelligent robots and/or avatars it is important to consider human psychology, human socialisation, and the way humans relate and interact; as well as consider the theory of attachment.

In conducting this initial literature review, I considered Schermer's (2001) statement as an acknowledgement of the 'lens' through which I explored literature relating to human psychology, socialisation, and interaction; namely that:

> while psychoanalysis, like the humanities, is introspective, intersubjective, and interpretive, it is consilient with the natural sciences, … First and foremost, the constant aim of psychoanalysis is to contribute to an understanding of human experience that is an integral part of a scientific world view … Moreover, …, psychoanalysis often taps neuroscience … for insights that might link it to biology. Psychoanalytic interpretations, while they do not themselves meet the criteria for scientific proof, both incorporate and supplement the "story" of human nature and evolution as understood by the sciences. (pp. 22-23)

Relational psychoanalysis grew from the object relations theoretical perspective (Mitchell & Greenberg, 1983), with Mitchell and Black (1995) asserting that object relations theory reflected the new focus where the "fundamental motivational push in human experience is not gratification and tension reduction, using others as a means toward that end, but connections with others as an end in itself" (p. 115).

In investigating human beings' experiences interacting with AI social systems, through a psychotherapeutic lens, the above explanations are considered when exploring human relational experiences, connections, and attachment. Cundy (2015) considered that some people may willingly accept or oppose the emergent digital culture and wondered how attachment might impact aspects of human beings' willingness to accept these AI social systems. The enmeshment and merging of the real and virtual worlds, is suggestive of a merging or lack of distinction between real and virtual relating, and a curiosity around how human beings experienced their interactions with AI robots and/or avatars. Neufeld (2015), a developmental psychologist with an attachment theory perspective, stimulated my thinking and questioning in relation to human experiences with AI robots and avatars in an interactional context and how it may inform the future. Neufeld and Mate (2013) stated that "technological innovations should be understood, not in terms of their content but in terms of how they change society. When we create a new technology, we are changing ourselves in fundamental ways" (p. 295). To provide a framework and broader thinking around how human beings relate, connect, and socialise, I included in my initial literature review some key attachment theorist authors, such as Holmes (2001), Bowlby (1969), Neufeld (2015), and Neufeld and Mate (2013). For example, Holmes (2001) stated that childhood attachment relationships significantly shape and effect relational abilities and their mental representations as adults. Attachment can be seen in terms of comprehending human development and ways of relating in regards to the notion of the dyad of 'self-other' (Cundy, 2015).

When exploring attachment theory, I noted Holmes' (2001) commented that human beings safety and wellbeing, both physically and emotionally, relies entirely on social bonds with other human beings. This security begins with infants, who can become securely or insecurely attached, with three main categories of insecure attachment: avoidant, ambivalent, and disorganised. "Secure attachment in childhood is reflected in adult life by the ways in which people talk about their lives, their past and in particular their relationships and associated mental pain" (Holmes, 2001, p. 7). Cundy (2015) viewed having secure attachment in relating to others implies trust, warmth, and openness, a greater capacity to nurture and be nurtured, along with more autonomy—we explore further and more comfortably from a secure base. Cundy (2015) referred to Winnicott's suggestion that these nurturing maternal functions are crucial in providing the context for 'indwelling' where the infant's mind becomes established in their body which results in the self being "experienced as an embodied entity" (p. 6).

In comparing the differences in how infants attach and their future development and ability to relate and connect to themselves and others throughout their lives, Cundy stated that "with insecure attachment, arising when an early significant relationship has been insufficiently secure and containing (and not remedied by later secure relationship), connecting with others will be more of a challenge" (p. 194). Cundy (2015) referenced Winnicott's' notion that "an infant who is not securely held…grows with the sense of himself as lifeless, fragmented, disembodied, or depersonalised. He is not safely contained within a psychic skin" (pp. 6-7). Thus, indicating the importance of an infant having a secure attachment and how it can present as an adult in subsequent interactions and relationships. When a human baby is born, its focus is on its need for survival. It learns and develops according to its environment, including the other humans, and how they relate to each other and to the infant. Survival is also influenced by the larger society and the culture within which those people find themselves and impacts communication. There is an innate human need for children to be in relationship (i.e., emotionally attached) with their parents or caregivers and with their community. Ainsworth (1969) highlighted that:

> attachment refers to an affectional tie that one person (or animal) forms to another specific individual...Once formed…an attachment tends to endure... This implies the formation of intra-organismic structures, presumably neurophysiological in nature, which provide the person with a continuing propensity to direct [their] attachment behaviours toward specific objects of attachment. (pp. 2-3)

When considering human relational experiences, connections, and attachment theory, it is essential to acknowledge the impact of a person's social context and real events, as part of their relational milieu, on their being; as our reality is determined by meanings in

relation to time, context, culture and are value-bound (Bowlby, 1993). This is reflected by Cundy (2015) who considered that attachment theory focused on the impact of the external world on internal reality. Carlino (2011), a psychoanalyst, stated that "socio-cultural changes and technological innovations that settle into and circulate within society affect both reality and people's subjectivity" (p. 1). Cundy (2015) explained that:

> we are primarily such social creatures, we human beings, that we roll, sloosh, wallow, and emerge within our connections with others, whether they are on- or offline. We have the capacity to thrive within the social milieu, whatever milieu it is, as long as we are able to engage with others wholeheartedly. (p. 195)

Neufeld (2022) stated that attachment is key to survival; the drive or preeminent need for togetherness and to increase the probability of survival in mammals, as part of an evolutionary nature. Togetherness is linked to the emotional system of caring. These strands of connection—to care for, receive care, and care about (i.e., attachment)—are not visible but hold human beings together (just as we do not see what holds the universe together).

According to Neufeld (2022) human beings are not meant to be conscious of attachment and love, as "knowledge is the lesser of the two evils"; hence the reason for shining light on this topic in the context of the current study. Neufeld reminded that Aristotle had observed how he and his fellow philosophers wanted to spend time together, a sign of their emotional attachment to each other. However, Neufeld pointed out a significant flaw in the social construct; namely, that the attachment has been to peers (seeking peers to meet social needs) and not to the attachment to family and village through the constructs of hierarchy. As a result of World War II, and the 'fatherless society' (Neufeld, 2022) that followed, there was not a sense of cascading care (vertical, hierarchical relationships); rather, peer-oriented attachment (horizontal or vertical relationships), projecting unsafe extensions of ourselves and the environment.

Neufeld (2022) noted that there are universal principles in attachment (c.f., Ainsworth et al., 1978; Bowlby, 1969/1982, 1973, 1980,1988), which change the way one attaches. The first attachment is through the five senses (i.e., smell, taste, touch, sight, sound). Then moving from need to be with, to needing to be the same; then to be part of or belong to, then needing to be important or belong to, understanding physical and emotional needs are different. As such, the developmental destiny is the source of fulfilment to cascading attachment with family members. Therefore, Neufeld stated that if a child does not go into the world with a sense of their own values and societal values, the result is a feeling of lack of connection and togetherness. As stated by Bowlby (1969/1982) "attachment behaviour does not disappear with childhood but persists throughout life" (p. 350).

The concept of attachment has remained as relevant today as when Bowlby wrote about it. The preeminent need for attachment to others has not changed; however, the tools and ways of connecting and communicating have. Stadter's (2013) commented that it is not only the functions of technology that are pertinent to study but also how this technology affects us, which resonated further with this study asking how human beings' experience AI social systems in an interactional context.

Balick (2014) noted that the central drive for using social technologies (including AI robots and avatars) is because it is relational in nature; "other social technologies … have become the new technologies of our intimacies" (p. xxviii). Attachment is an innate need in all humans with the core and leading need is attachment to others (Neufeld & Mate, 2013) and connection which is the basis for the survival of the human species.

Cundy (2015) referred to Balick's view that these technological opportunities can lead to human beings being, to some extent, compulsive in their need for this social technology, as it does not fully satiate and can be seductive, providing an alluring sample of a deeper, innate human desire. This human desire for togetherness, suggests humans are driven by a need to connect and socialise, whether with other humans or technology. However, the implication is that technology is unable (to date) to 'entirely satisfy' the human need for connection; rather, social technology substitutes for 'real' human connections. Attachment, according to Neufeld and Mate (2013), "is the key to explaining the shape the digital revolution has assumed" (p. 268). They noted that no matter whether people interact with each other in person, telephone, or another technological medium, attachment, more than any other element, influences human communication and relating.

Turkle (2011) suggested that "these days, insecure in our relationships and anxious about intimacy, we look to technology for ways to be in relationships and protect ourselves from them at the same time" (p. xii). This sentiment is reflected in Cundy's (2015) suggestion that attachment styles influence how we use communication and social technology; thus influencing, and most likely reinforcing, attachment experiences. It also highlights how "the virtual world both enables and obstructs a variety of ways of relating" (Balick, 2012, p. 122). While our preeminent need for attachment to significant others remains central to humans, Cundy contended that technology has exponentially advanced. Resulting in an ever-changing social setting, that include different ways of communicating, new symbols and use of language and new representations for thinking about ourselves.

Summary

The initial literature review explored academic research journals, books, conferences, and popular media, as the topic is much talked about from different avenues, whether business, education, innovation, or others. It included literature discussing the development of AI and human interactions with AI, and explored various areas of literature regarding AI technology, with a focus on implications of social AI, such as AI social systems that are designed for the purpose of interacting with people (Bartneck et al., 2020; Carpenter, 2013; Coulson et al., 2018; Goodrich & Schultz, 2007; Kanda & Ishiguro, 2013; Law et al., 2022; Parasuraman et al., 2001; Rabb et al., 2022; You & Robert, 2017). While this literature is beginning to consider 'the human being' in the AI social system interaction, a gap remains when considering human beings' experiences interacting with AI robots and/or avatars. Bowlby (1969/1982) stated that in "the satisfactory development of attachment …there is [an] urgent need to be able to distinguish favourable development from unfavourable and also to know what conditions promote the one or the other" (p. 331). This statement seems just as relevant today; however, now the question could be: in our contemporary times of the technological era what might the current conditions be for relating and connection? Furthermore, statements from authors, such as Turkle (2011), who posited that "technology proposes itself as the architect of our intimacies" (p. 1), reflect the need for further research; thus, stimulating the current study, with the focus around the technological advances in AI robots and avatars and human beings' experience interacting with them.

Harari (2018) and Turkle (2011) each put forward a question that lingered with me throughout the research process: "what will safeguard [human beings'] physical survival and their psychological well-being? We cannot wait for the crisis to erupt in full force before we start looking for answers" (Harari, 2018, p. 37) and "in all of this, there is a nagging question: Does virtual intimacy degrade our experience of the other kind and, indeed, of all encounters, of any kind? (Turkle, 2011, p.11). This preliminary literature review ascertained the potential depth and quality of data available to support the study. Consequently, it became evident that while there is literature on the topic of HRIs, digital conversational agents, and some research and discussion papers around the area of AI and attachment, there appeared to be limited literature around the topic. Rather than attempting to investigate from an external/objective perspective, the current study was conducted from a subjective, heuristic perspective of people's interactional experiences themselves, and co-constructed with myself, as the researcher.

[According to John Dewey] Pragmatism considered the future and led to the conception of a universe whose evolution is not finished, … which is still, … 'in the making', 'in the process of becoming', of a universe up to a certain point still plastic. (Charmaz, 2014, p. 263)

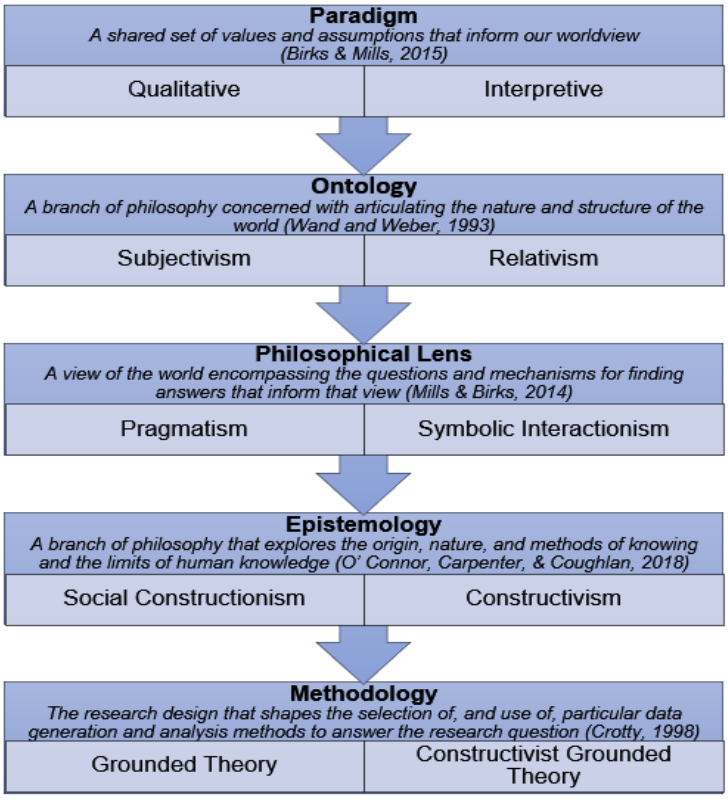
Chapter 3: Methodology

Introduction

This chapter discusses the constructivist grounded theory methodology used in the study. First, I provide an overview of my paradigmatic position (qualitative and interpretive). I discuss the central ontological (subjectivism and interpretivism) foundations, followed by an outline of my theoretical perspective and the philosophical lens used in the study. I outline the key epistemological underpinnings (social constructionism and constructivism) that contribute to the choice of grounded theory methodology. Next, I discuss the evolvement of grounded theory methodology and make specific reference to constructivist grounded theory, explaining the applicability of this methodology for the research and topic. An overview of the perspectives chosen for this study can be found in Figure 3.1. The chapter concludes by reviewing the grounded theory implemented in this study (explained in more detail in Methods Chapters 4 and 5).

Figure 3.1.

Perspectives Chosen for This Study



Paradigmatic Positioning

A paradigm is “a shared set of values and assumptions that inform how we view the world” (Birks & Mills, 2019, p. 4), that correlates with a methodological perspective and method (Olsen et al., 1992). Qualitative researchers are interested in the study of social and cultural phenomena using an interpretive, naturalistic approach to explore, represent, and understand the complexities of people and their social and cultural contexts (Myers, 2009; Philip, 1998). Sources of qualitative data can include observation and participant observation, interviews and questionnaires, documents and texts, and the researcher’s impressions and reactions (Myers, 2009). When selecting the research paradigm, I reflected on Guba’s (1981) suggestion that “it is proper to select that paradigm whose assumptions are best met by phenomenon being investigated” (p. 76). I selected a qualitative research paradigm as I aimed to explore determinants and social processes that highlighted specific phenomena (Clarke et al., 2015); namely, human beings’ experiences interacting with AI robots and/or avatars. As such, “the goal of research is to understand the complexity of the human experience” (Loseke, 2013, p. 22).

In determining my paradigmatic position, I concur with Denzin and Lincoln’s (2005) statement that “all research is interpretive; it is guided by the researcher’s set of beliefs and feelings about the world and how it should be understood and studied” (p. 22). The current study can be said to follow an interpretive approach to the people being studied and the meaning they attribute to the studied phenomena. My interpretive position, as researcher, assumes that knowledge and meaning are interpretations derived through social constructions such as language, consciousness, and shared meanings; and are located in a particular context, re-interpretated, and negotiated through dialogue (Crabtree, 2006). Charmaz (2014) acknowledged grounded theory as an interpretive tradition where researchers reflect externally and internally when constructing theory from various thoughts and experiences; contending that we are “part of our constructed theory and this theory reflects the vantage points inherent in our varied experience, whether or not we are aware of them” (p. 260).

The aim of interpretive theories is to understand how people construct meanings and actions; thus, bringing in the subjectivity of the human being and acknowledging the subjectivity of the researcher. These interpretive theories assume “emergent, multiple realities; truth is provisional; and social life as processual” (Charmaz, 2014, p. 231) and include, amongst others, symbolic interactionist and social constructionist perspectives. Consequently, George Herbert Mead’s pragmatist philosophical position is consistent with the interpretive view and informs symbolic interactionism.

Mead believed in "a sophisticated view of action as the starting place for analysis that includes the person's imagined understanding of the other person's role and response during interaction" (Charmaz, 2014, p. 231). Hence, I aimed to interpret my participants' meanings and actions and they interpreted mine. My paradigmatic position is reflected in Charmaz's (2014) summary:

> Interpretive theory aims to: Conceptualize the studied phenomenon to understand it in abstract terms; Articulate theoretical claims pertaining to scope, depth, power, and relevance of a given analysis; Acknowledge subjectivity in theorizing and hence recognize the role of experience, standpoints, and interactions, including one's own; Offer an imaginative theoretical interpretation that makes sense of the studied phenomenon. (p. 231)

The interpretivist paradigm puts forward the notion that the researcher's values are intrinsic in all phases of the research process (Crabtree, 2006). For Alasuutari, theories present interpretive frameworks through which to view realities. Constructivist grounded theorists follow a similar perspective, by "build[ing] from specifics and move to general statements while situating them in the context of their construction" (Charmaz, 2014, p. 232). In other words, the interpretivist paradigm emphasises placing analysis in context and understanding the subjective experiences of the people being studied; thus, seeking to explain the subjective meanings that effect the social action. The interpretive perspective views reality as socially constructed and fluid, where what is known is continually negotiated and re-interpreted within cultures, social situations, and relationships with other people (Crabtree, 2006).

Interpretivism posits that "humans construct knowledge as they interpret their experiences of and in the world; rejecting the objectivist notion that knowledge is simply there to be identified and collected" (Hiller, 2016, p. 13); thus, viewing all knowledge as grounded in our specific experiences, which are subjective and linked to the everyday contexts in which we live (Denzin & Lincoln, 2005; Greene, 2010). From an interpretivist perspective, I understand that participants have their own views and beliefs that are contextually and socio-culturally bound and acknowledge my own beliefs and worldview that I bring to the interpretive research process.

Ontology

"Ontology refers to a branch of philosophy concerned with articulating the nature and structure of the world" (Wand & Weber, 1993, p. 220). It is concerned with the form and nature of reality, and what can be known about it. Ontological assumptions guide the types of research questions asked in terms of understanding how the world operates or how people act or interact (Hiller, 2016).

As such, the ontological position considers the researcher's relationship with the reality of the study which, for me, is subjective; and that reality is socially constructed, in other words reality is inherently relational. Therefore, ontology, from the perspective of relativism, views what we know about reality as being intersubjectively constructed through meanings and insights as a result of social interactions (Crabtree, 2006).

The ontology of relativism is that there is no universal, objective truth; rather, "standards of reasoning, and procedures of justification are products of differing conventions and frameworks of assessment and that their authority is confined to the context giving rise to them" (Stanford Encyclopaedia of Philosophy, 2015). As such, the ontological view of relativism considers that views, aspirations, experiences, and behaviours are within the context of cultural assumptions, interests, and beliefs. Given the ontological position of relativism, the current study aimed to conceptualise descriptions and meanings of human beings' experiences interacting with AI social systems, within their subjective context.

Charmaz (1991) defined the 'self' as "the organised set of internalised attachments, commitments, attributes, images and identifications with which a person creates a concept of self" (p. 72); a position that correlates with my worldview and definition of 'self'. Reflecting on my ontological position, I consider that we all have our own cognitive and cultural biases which shape our perceptions and senses and form part of our individual and social subjectivity. As the researcher, I too hold presuppositions. Subjectivism views social phenomena as being created from the perceptions and actions of the social actors who form a part of their reality (Saunders, et al., 2009).

Theoretical Perspective: Pragmatism and Symbolic Interactionism

Mills and Birks (2014) defined philosophy as "a view of the world encompassing the questions and mechanisms for finding answers that inform that view" (p. 18). Researchers consider their worldview, philosophical values, and interpretation before beginning research (Birks, 2014) through questions that may include "what do people assume is real? How do they construct and act on their view of reality? Thus, knowledge and theories are situated and located in particular positions, perspectives, and experiences" (Charmaz, 2000, p.127). These research decisions are situated in a theoretical perspective which explicates the researcher's position, essentially giving an overview of the philosophical assumptions that shape how the researcher integrates their views about the relationships between the types of knowledge that are sought and what the process is to achieve it (Hiller, 2016). Noting my values and assumptions as intrinsic/evident throughout each stage of the research process equipped me to make methodological decisions, influenced by my beliefs about the nature of reality.

This in turn, determined how the qualitative methods (i.e., essential grounded theory methods) were used (Birks & Mills, 2011, 2015; Fendt & Sachs 2008; Howell, 2013). In considering my philosophical position, I draw on Mitchell and Black (1995) who posited that the further development of the object relations tradition by Winnitcott, Klein, Fairbairn, and others; and development of relational analysis, has led to object relations theory reflecting a new focus in which the "fundamental motivational push in human experience is not gratification and tension reduction, using others as a means toward that end, but connections with others as an end in itself" (Mitchell & Greenberg, 1983, p. 115).

Consistent with constructivist philosophy, research does not occur in a vacuum; instead, is shaped and informed by the researcher's perspective in which they are operating (Charmaz, 2014; Kenny & Fourie, 2015). I aligned with Charmaz' (2014) constructivist grounded theory approach, considering the 'lens' through which my tacit knowledge and theoretical perspectives informed my philosophical understandings about the nature of being in the world and knowledge. Howell (2013) highlighted that for constructivist grounded theorists, "knowledge, truth, reality and theory are considered contingent and based on human perception and experience" (p. 16). By explicating my philosophical position and reflected chosen methodology, theoretical perspective, ontology, and epistemology, the aim has been to enhance the rigour and credibility of the study (Birks & Mills, 2015; Ward, Hoare, et al., 2015). My theoretical perspectives for this study are pragmatism and symbolic interactionism.

Pragmatism

The pragmatist philosophy was developed by Charles Sanders Peirce and John Dewey. It was core to both social work and the philosophy of the Department of Sociology, University of Chicago (Locke, 2007; Strübing, 2007), and focused on human behaviour and interactions in relation to a wider society (Charmaz, 2014). Pragmatism is a philosophical movement that claims that "reality does not exist independent of meanings that are created, defined and acted upon by people according to their usefulness [of the conceptual or theoretical outcome]" (Robbins et al., 2006, p. 321). As such, pragmatism has been defined as a 'philosophy of action', where people engage in an ongoing process of practice, inquiry, and adaptation in an ever-changing social world (Corbin & Strauss, 2015), and develop meanings through useful actions to solve problems (Charmaz, 2014; Strübing, 2007). Dewey's action-oriented theory of knowledge considered the close connections between knowledge and everyday actions, and the related consequences (Biesta, 2010; Corbin & Strauss, 2015; Jeon, 2004).

Being influenced by Darwin's theory of natural selection, a naturalistic account regarding species and the interrelationships between organisms and environments, Dewey rejected the dualistic epistemology and metaphysics of modern philosophy, favouring a naturalistic approach that considered knowledge to be the result of an active restructuring/adaptation of people to their environment and conditions (Field, n.d.). Additionally, Dewey (1938a) considered the value of any experience to be linked with past, present, and future; a lifelong evolving process where knowledge has a practical instrumentality in guiding, understanding, and directing that interaction, and subsequent situations. Dewey purported 'the experimental theory of knowledge' which considers all knowledge to be provisional, being determined based on how useful it is for people or a particular situation (Bryant, 2009).

Another line of pragmatism considers meaning, or the truth of an idea, results from its possible consequences; and that the consequences of human beings' constructions of meaning should be ongoingly assessed in order to determine if they contribute to human well-being (i.e., usefulness) (Hickman et al., 2009). Pragmatist assumptions consider the benefit/usefulness of theories or beliefs is dependent on effective practical application, as meanings emerge through practical actions to solve problems. In other words, through actions human beings come to know the world. Thus, pragmatism considers "scientific truth is relative, provisional, and assessed through what works in empirical practice" (Charmaz, 2014, p. 263). Indeed, the early pragmatists—Peirce, Dewey, and James—contested the principle of absolute truth, with Dewey disaffirming what he referred to as "the spectator theory of knowledge" (Bryant, 2009, p. 45), implying objective and passive observation. Pragmatists, James and Dewey, saw people's experiences, actions, and interactions as interlinked in their social and physical environments, determining how they come to know their world (Bryant, 2009; Charmaz, 2009, 2014; Corbin & Strauss, 2015). Consequently, the environment guides and limits human action in specific ways and is fluid, resulting in reality being open to multiple interpretations (Charmaz, 2009, 2014). Pragmatism is concerned with the developing of ideas on the world as research is ongoing, highlighted by the aim of 'knowing' rather than knowledge itself. Additionally, pragmatism holds that researchers need to be cognizant of their own role in the research process, considering attributes such as positionality, orientation, diversity, and reflexivity (Bryant, 2009).

Many disciplines embraced pragmatist assumptions, including psychology and social psychology. Strauss, who studied at the University of Chicago, within the symbolic interactionist and qualitative approaches of inquiry, "brought [with him] the pragmatist philosophical study of process, action, and meaning into empirical enquiry through grounded theory" (Charmaz, 2000, p. 512).

Grounded theory is a research methodology that has its foundation in pragmatist and symbolic interactionist philosophies (Stern & Porr, 2011) and interprets human action in the context of problematic situations in a shifting world, in order to comprehend what is problematic, what is of value to people, and the methods implemented to facilitate change and to reach an outcome (Hunter & Krantz, 2010).

Pragmatism and Grounded Theory.

Grounded theory methods have their epistemological roots in pragmatism and symbolic interactionism. Bryant (2017) highlighted that *"*what GTM [grounded theory method] and pragmatism have in common is a concern with people's engagement with the world, reliant on detailed observation and insight, followed by never-ending and iterative efforts to comprehend, persuade, and enhance" (p. 346). Pragmatism is epistemologically aligned with constructivist grounded theory, where both the pragmatist approach and constructivist perspectives emphasise action and meaning (Charmaz, 2014). Constructivist grounded theory, rooted in pragmatism, preserves an emphasis on actions, meanings, and language, and encourages an 'interpretive rendering' of the studied world rather than an objective account of events and reports (Charmaz, 2014). Charmaz (2003, 2006, 2009) developed a constructivist version of grounded theory that aims "to build data gathered from the phenomena of the study, to explain the processes which occur within the phenomena. It seeks to capture the complexity of actions/strategies that are employed within the phenomena of study" (Smythe, 2012, p. 5). Grounded theory is described as an iterative and circular process in which include three types of reasoning are used for analysis and theory construction: induction, deduction, and abduction. Abduction was a concept developed by Peirce (1878, 1958) who identified limits to inductive reasoning and sought new theoretical explanations to account for surprising data that emerged during the inductive process (Charmaz, 2014). Abductive reasoning in grounded theory developed from the pragmatist perspective of problem-solving. Richardson and Kramer (2006) clarified the link between abduction and pragmatism, stating that "abduction is the process by which useful explanations are developed and is therefore an essential concept within pragmatism. This process of finding useful explanations is essentially 'an inference' from observed fact" (p. 498). Mead's concept of emergence acknowledges that the present and past reality differ and that new influences or factors in the present experience led to new interpretations and actions (Charmaz, 2014).

Symbolic Interactionism

Mead integrated notions developed by other sociologists of the time (e.g., Dewey, James, Thomas, Znaniecki, and Cooley), into a single theory called symbolic interactionism (Blumer, 1969; Crotty, 1998; Mead, 1934; Strauss, 1977).

In '*Mind, Self and Society'*, Mead (1934) outlined his theory of the development of the social self, describing the "human [as a] biological organism [that] possesses a mind and a self" (as cited by Herman-Kinney & Verschaeve, 2003, p. 214).His ideas formed the foundation of what Blumer termed 'symbolic interactionism' (Charmaz, 2014). Blumer (1969) hugely contributed to the establishment of symbolic interactionism as a sociological social psychology. Mead viewed the symbolic interactionist tradition as consistent with interpretive theory, as he noted that researchers interpret their participants' meanings and action as well as their own (Charmaz, 2006). Symbolic interactionism was thus informed by the pragmatist tradition.

Symbolic interactionism views meanings as social outcomes, shaped in the context of social interactions "formed in and through the defining activities of people as they interact" (Blumer, 1969, p. 5). The pragmatist approach imbued symbolic interactionism with a focus on human agency, language, and interpretation; hence, symbolic interactionists highlight the importance of language on self and society and understanding the subjective meanings within human worlds emerging out of social interaction and experience (Charmaz, 2014; Fisher & Strauss, 1979). Furthermore, this perspective assumes that people construct selves, society, and reality through interaction, thereby creating a dynamic process of meanings and actions that influence each other. Thus, it is assumed that people are dynamic, creative, and reflective; and that social life involves processes (Charmaz, 2006, 2014). Symbolic interactionism is a perspective that includes an abstract theoretical framework of assumptions for viewing social realities which provides a way of enhancing the knowing and developing of meanings, actions, and situations in the studied world. "Symbolic interactionism is a constructionist perspective because it assumes that meanings and obdurate realities are the product of collective processes" (Charmaz, 2006, p. 189). Pragmatism informs symbolic interactionism as it considers human action and interaction in the construction of self, reality, and society; where language and symbols are assumed to contribute a central part of constructing and sharing meanings and actions (Charmaz, 2014). This perspective "views interpretation and action as reciprocal processes, each affecting the other" (Charmaz, 2014, p. 262); thus, recognising that individual and collective actions and meanings are consequential.

Charmaz (2014) identified three reasons for symbolic interactionism to have its foundations in pragmatism. First, grounded theorists begin their research from an open-ended theoretical perspective that is rooted in the pragmatist tradition. Second, the emphasis in pragmatism on process and change corresponds with grounded theory methods, sensitising researchers to reflect on change in novel ways.

The perspective of emergence recognises that past and present reality/experiences can differ and give rise to new interpretations and actions. Third, grounded theory (especially Straussian and constructivist versions) uses abductive reasoning. Crotty (1998) highlighted the links between pragmatism and symbolic interactionism when he stated that “symbolic interactionism is pragmatism in sociological attire” (p. 62). Symbolic interactionism is a key theoretical perspective associated with grounded theory. It was Strauss who included the reasoning and assumptions of symbolic interactionism into the methods of grounded theory (Charmaz, 2014). Symbolic interactionism stimulates theoretically driven research, leading to potentially new areas of analysis, ideas, and perspectives of the studied phenomenon. Symbolic interactionism can guide the researcher regarding what is described about the data and can inform how the researcher may come to understand and conceptualise how human beings’ experiences have been shaped and the subjective meanings they give to their experiences with AI social systems in the 21st century (Aldiabat & Le Navenec, 2011).

Symbolic Interactionism and Grounded Theory.

The way people socialise is primarily through interaction. Children come to know the world around them by interacting with adults and socialising with others. These interactions with adult attachment figures, usually parents or caregivers, provide children with their self-image and worldview, and are core to the evolutionary development of a child’s perception of the world. Further, these interactions influence the child’s behaviour, actions, and emotional world; and, subsequently, how they perceive and experience reality (Gabatz et al., 2017)**.**

Consequently, the self-concept and the learning of symbols occurs out of interacting with other people, first by observing others’ behaviours, concepts, and meanings which reflect the child’s perception of self. Charmaz (2014) highlighted how human beings perceive and act differently, thus behaving in situations and towards things based on the meaning they have for them. Like the pragmatist approach, people are viewed as dynamic, creative, and reflective, and are continuously involved in a process of interpreting and defining while acting in response to their view of each situation (Eaves, 2001). Symbolic interactionism posits that “social reality and society should be understood from the actors who interpret their world through and social interaction” (Van Manen, 1990, p. 186). Therefore, the grounded theory method can be seen as a useful way to study social interactions, discourse, and experience, as it aims to explain a process. Significant topics in symbolic interactionism include meaning and the concepts of self, action, and interaction. These interactions depend largely on the use of symbols; hence, the term ‘symbolic interactionism’ (Charon, 2010).

Once humans develop meanings for objects, they then go onto to use the objects in their interactions; therefore, symbolic interactionism speaks to the complexity of human relating and the interaction and experiences with AI robots and/or avatars.

Central to symbolic interactionism are human beings' social interactions with each other, where our expressions and movements are interpreted by each other, and our responses and reactions are based on these interpretations. Consequently, when in social interactions we tailor our behaviour in relation to the behaviour of others which involves both benefits and costs that can then result in a positive or negative impact on human well-being (Aldiabat & Le Navenec, 2011).

The central goal of grounded theory is consistent with the overall goal of symbolic interactionism, in that grounded theory provides a hypothesis for exploring human behaviour (Chenitz & Swanson, 1986). Glaser (1978) stated that the central goal of the grounded theory method is to discover the basic social process(es); namely, "the theoretical reflections and summarizations of the patterned, systematic uniformity flows of social life which people go through, and which can be conceptually 'captured' and further understood through the construction of basic social process theories" (p. 100). As a qualitative researcher, I concur with Aldiabat and Le Navenec (2011) who advocated for the compatibility of symbolic interactionism and grounded theory methodology for studying human behaviour because:

> Symbolic Interactionism provides a guiding framework to collect data about the meaning of a particular type of behaviour and the contextual sources of such meanings, and how they change in and through social and physical time and space; and (b) Grounded Theory methodology affords a systematic approach to generate a theory that illuminates human behaviour as a social process among actors in their interactional context. (p. 1068)

Additionally, Charmaz (2014) viewed symbolic interactionism and grounded methods as a suitable fit, that inform and advance each other, referring to them as a "theory-method package", suggesting "symbolic interactionism offers a worldview and language for conducting grounded theory studies" (p. 284). For Charmaz, as a result of grounded theory and, to some extent, symbolic interactionism being applicable across disciplines and professions, the "theory-package of symbolic interactionist-grounded theory" (p. 281) is attractive beyond sociology and microsociology. Adele E. Clarke (2005) also advocated that grounded theory and symbolic interactionism comprise a theory-method package in which ontology and epistemology are 'co-constitutive' and unique. In other words, symbolic interactionism can guide analysis of commonplace experiences and practices and grounded theory provides the researcher with the tools to gain theoretical insights from these experiences and practices (Charmaz, 2014).

Using the symbolic interactionist perspective and focusing on meaning extends to studying emotions among participants, while being aware of the researcher's own emotions and how they enter both research relationships and interpretations (Charmaz, 2014). Thus, Charmaz (2014) claimed that a symbolic interactionist perspective enhances the degree of reflexivity to which constructivist grounded theorists aim. Further, Charmaz (2014) stated that symbolic interactionism offers "an open-ended perspective than can inform grounded theories" (p. 277). Blumer (1969) asserted that symbolic interactionism is based on four points: First, human beings (individually and collectively) act toward things in their environment on the basis of the meanings that those things have for them. Second, these meanings derive from, or arise out of, the social interaction (communication) between and among people (interpreting each other's signals). Communication is symbolic because we communicate via languages and other symbols. Third, these social acts and meanings are recognised and modified through an interpretive process wherein "the actor selects, checks, suspends, regroups, and transforms the meanings in light of the situation in which he is placed and the direction of his action… meanings are used and revised as instruments for the guidance and formation of action" (Blumer, 1969, p.50). The fourth point refers to "the complex interlinkages of acts that comprise organisation, institutions, division of labour, and networks of interdependency are moving and not static affairs" (Blumer, 1969, p. 50).

Charmaz (1980, 2000, 2002, 2014) reformulated and clarified Blumer's premises emphasising, from a constructivist grounded theory approach, that in symbolic interactionism meanings are interpreted through reciprocal language and communication; addressing the social interactions in which people make and mediate meanings and reality through a continuously emerging methodological process. Charmaz (2000) interpreted Blumer's position by purporting that reality refers "to a world made real in the minds and through the words and actions of its members" (p. 523).

Symbolic interactionists take on a relativist position "where reality is constantly reformulating as a fluid construction of individuals and, in turn, their social reference groups" (Birks & Mills, 2015, p. 51). Similar to most symbolic interactionists, Charmaz (2014) held the view that it is not possible to separate the researcher from the participant in the generation of data and acknowledged that the empathetic understanding of the participants' own communicated meanings, actions, and worlds is important. Constructivist grounded theorists consider that empathetic understanding assists in advancing the study beyond a descriptive version towards constructing an interpretive understanding of the world. Symbolic interactionism is a theoretical perspective that acknowledges the existence of social structures and argues that people create and replicate them.

As such, Blumer (1969) highlighted the importance of interpretation in that interpretation and action arise from interaction where the reconstructed past, lived present, or imagined future is attended to (Charmaz, 2014). Thus, it is understood that interaction is a symbolic process, dependent on spoken and unspoken shared language and meanings occurring within a socio-cultural and historical context that influences but does not define it (Blumer, 1960). Mead (1934) posited that language is central for the development of self and human behaviour in a social world.

Charmaz (2014) noted that Mead saw the self and mind developing at the same time. Cooley's (1902) idea of the 'looking-glass self' holds relevancy for this study, as it describes how people receive reflections from mirrors similar to how we imagine others perceive us, and we react to these reflections by experiencing feelings, such as pride or shame. Thus, people shaping their responses to these imagined perceptions, assess its meaning and influence any responses that may result in an action (Charmaz, 2014). Charmaz stated that:

> symbolic interactionists view the self both as a continual unfolding process and as a more stable object, the self-concept, or ... 'real' self, [contending] that a person's habitual ways of being and doing can bridge the fluidity of self as process and the self as object. (p. 268)

As such, self as process and the reconstruction of self and identity result out of new modes of being in the world. In summary, symbolic interactionism and grounded theory can be seen to provide a strong theory-method package, with each informing and advancing the other. Symbolic interactionism gives a perspective and language for conducting grounded theory research. Charmaz (2014) highlighted that in symbolic interactionism the focus on meaning and definition includes investigating participants' emotions; while researchers also need to be aware of their own emotions, and how these impact their research relationships and interpretations. Additionally, symbolic interactionism advances the type of reflexivity to which grounded theorists aim (Charmaz, 2014). Denzin (1989) described symbolic interactionists as seeking to understand the social world being studied, with theory being interpretive, grounded, and remaining close to the data.

Epistemology: Constructivism and Social Constructionism

Epistemology is branch of philosophy that investigates the source, dynamics, and methods of knowing and the limits of human knowledge. Grant and Giddings (2002) defined epistemology as "the nature of the relationship between enquirer and known, what counts as knowledge, and on what basis we can make knowledge claims" (p. 12).

Accordingly, a researcher's epistemological position substantially effects their choice and use of research methodology, acting as a 'lens' though which the research topic is approached (Kelly, 2010; Lipscomb, 2008; O' Connor et al., 2018). Epistemological beliefs are echoed in the theoretical perspectives, methodology, and methods of the study, where "depending on a researcher's beliefs about what can be known (ontology) and how to approach coming to know it (epistemology), different decisions will be made toward designing an effective study" (Hiller, 2016, pp. 99-100). When considering the epistemological assumptions of constructivist grounded theory, I reflected on Charmaz' (2014) statement that all knowledge is constructed, and that the reality of the studied world is fluid and subject to changes based on research participants' subjective experiences and constructions of reality. There are close correlations between pragmatism and constructivism, with early pragmatists (e.g., Peirce, Dewey, James) acknowledged as predecessors of constructivist notions (Reich, 2009). Here, I consider the epistemological positions of constructivism and social constructionism. Charmaz (2006) defined constructivism as

> a social scientific perspective that addresses how realities are made. This perspective assumes that people, including researchers, construct the realities in which they participate. Constructivist inquiry starts with the experience and asks how members construct it. To the best of their ability, constructivists enter the phenomenon, gain multiple views of it, and locate it in its web of connections and constraints. Constructivists acknowledge that their interpretation of the studied phenomenon is itself a construction. (p. 187)

A constructivist perspective emphasises the importance of the studied phenomena, and views both data and analysis as constructed from shared experiences and relationships with participant relationships (Charmaz, 1990, 2001, 2006; Charmaz & Mitchell, 1996). In other words, constructivism considers that researchers do not discover knowledge; rather, knowledge is constructed where concepts and ideas are created to understand and interpret experience, and these constructions are continually tested and adjusted in considering new experiences. From a constructivist stance, realities are multiple with a resulting multiplicity of perspectives. The researcher is 'embedded' in their own worlds as opposed to separate from them, and potentially has multiple views which may conflict with research participants' views and realities.

Constructivism, a term used in social sciences, evolved from developmental and educational psychology, in particular a learning theory formed by Piaget (1964) and Vygotsky (1978) which was concerned with childhood development stages where knowledge is individually constructed through experience and social relationship (Ward et al., 2015; Young & Collin, 2004).

The epistemological stance of constructivism "acknowledges multiple, socially constructed truths, perspectives and realities" (Hunter & Krantz, 2010, p. 208). Central to constructivism is the view that people actively construct, rather than discover, their own personal knowledge, meaning, and reality in relation to their subjective experiences (Schwandt, 2000). Similar to pragmatism, constructivism offers a foundation for generating practical, useful solutions to problems, and meaning relevant to people, rather than constructions for the sake of constructions (Reich, 2009). Constructivism is viewed as an ontologically relativist and subjectivist perspective (Charmaz & Bryant, 2010).

Relativism is evident in the constructivist approach when, given a framework of shared understandings and even in regard to the same phenomenon, individuals may construct distinct meaning in different ways, leading to multiple realities and views which shift and modify in different situations (Charmaz, 2014; Charmaz & Bryant, 2010). Therefore, the constructivist perspective considers knowledge as subjectively influenced by individual constructions (Charmaz, 2014). As previously noted, knowledge and meaning are recognised as culturally and historically situated and contextually bound (Kuhn, 1970). With the study of human interactions and experiences, the focus is on understanding (Schwandt, 2003). As such, from a constructivist perspective, the meanings (and process of meaning making) that individuals attribute to their experiences are the most significant and are contextually bound. Guba and Lincoln (1994) considered constructivist research as epistemologically transactional and subjectivist by exploring meanings between the experiences of the research participant in their context, and those of the researcher. As such, constructivist researchers are concerned with both the experiences of the participants and the approaches they use to construct meanings from and in regard to their experiences. Alvesson and Sköldberg (2009) illustrated how human beings are always interpreting their experiences in some way; stating that "we never see single sense–data, but always interpreted data, data that are placed in a certain frame of reference or perspective" (p. 6). Thus, understanding that the meanings "ascribed from an individual's interpretations are relative—they belong to the knower in that time, place, and cultural situation. Knowledge and the knower are thereby inseparable" (Hiller, 2016, p. 111). Schwandt (2003) explicated constructivism as a form of perspectivism, referring to the specific lens (or conceptual framework) through which a person attributes the meaning of an experience or perception.

Further, the constructivist perspective views all insights as 'value-laden'; and there are many truths which can, at times, be distinct and conflicting (Pascale, 2011). Constructivism is concerned with understanding the meanings human participants attribute to their experiences of a specific phenomenon; as such, data generation methods involve providing a means for participants and researchers to communicate

their meanings and meaning-making processes. The use of language, between researcher and participants, is significant; thus, in constructivist methods, the use of linguistic processes of understanding is highly valued (Constantio, 2008). Accordingly, the use of transcriptions of open-ended interviews is a regularly applied method of data generation. Constructivism is an approach that asserts that meaningful human reality is not objective and is constructed by human beings through their interactions with others and their interpretations of the world (Alvesson & Sköldberg, 2009; Crotty, 1998). While Crotty (1998) commented, “there is no meaning without a mind” (pp. 8-9), it is important to note that meaning is more than what human beings create in and from their minds. Rather, meanings are constructions or re-constructions of individuals’ experiences of reality merging with insights within a context or situation that can change as a result of new experiences that shape new perceptions and knowledge (Schwandt, 2003). Sismondo (1993) and Andrews (2012) posited that, over time, the construction of individuals’ realities can result in them viewing these as objective realities, such as the ‘social structures of interaction with other human beings’ and assumptions as to how to act in the social world. An example could be if a person was to consider an AI social system to be their companion or friend, having constructed an individual reality as if it was an objective truth.

Situating the word ‘social’ in front of constructivism positions the epistemology in a social context and implies the way meaning is constructed (Crotty, 1998; Ward et al., 2015); thus, positing that social influences and experiences impact how people define their reality, and that knowledge is individually constructed (Dolitte, 2001). In other words, meaning is co-constructed within a social context where an individual’s reality is acknowledged, as well as understanding that often reality is constructed between people, through shared social experience and social engagement of meaning (Dolittle & Camp, 1999). Social constructivist approaches dynamically have altered many areas within the sociology field, including the development of new research linked to the sociology of the body, self, cultural identity, and systems of difference, which are clearly evident in the notion of embodiment being regarded as a symbolically facilitated process (Shotter 1993; Shotter & Gergen 1988). As such, constructivism is generally labelled by the term's social constructivism and social constructionism which have been used indistinguishably by different authors (Charmaz, 2003, 2008a, 2009; Sismondo, 1993; Ward et al., 2015). Thus, subject to the author and the context it applies, constructions may be knowledge, data, or objects (Sismondo, 1993). In grounded theory research, theories and interpretations are constructions based on data and observation (Ward et al., 2015). As stated above, by using the word ‘social’ in front of constructionism or constructivism, it positions the epistemology in a social context.

Constructivism does not completely align with socially constructed knowledge, as it has been defined as "individualist understanding of the constructionist position" (Crotty, 1998, p. 58). However, social constructionism highlights how culture, language, and social processes influence how people conceptualise reality and shapes our world view and how we feel and make meaning of our world (Crotty, 1998).

Charmaz and others have also referenced constructivist grounded theory as constructionist grounded theory (Bryant & Charmaz, 2007; Charmaz, 2008b, 2014; Ward et al., 2015), mainly to differentiate her version of grounded theory from the more objectivist versions of grounded theory and from the more extreme position of radical constructivism (Charmaz, 2008, 2014). Consequently, Charmaz (2008, 2014) specified that whether she uses the term 'constructivist' or 'constructionist' is dependent on the context of the discussion. Charmaz (2003, 2006, 2008, 2014) labelled her version of grounded theory constructivist, as she contended that the researcher is part of the research process and research outcome, which is a co-construction between researcher and participants.

Charmaz' version of constructivist grounded theory presents an approach to grounded theory that is socially and culturally bound, rejecting the concept of a 'tabula rasa', and recognising the necessity to shift the positioning of the research from objective observer to within a participant-researcher relationship (Ward et al., 2015). As a result of 20th century constructionism viewing research worlds as social constructions, but not as research practices, Charmaz (2014) wanted to differentiate her variant of grounded theory from the conventional social constructionist approach at that time (therefore, being more aligned with social constructivist approach), by calling her version of grounded theory constructivist grounded theory. Furthermore, by using the term 'constructivist' Charmaz (2014) recognised that subjectivity and the researcher's participation in the co-construction and interpretation of the data; stating that "subjectivity is inseparable from social existence" (p. 14). Subsequently, Charmaz (2014) has recognised the evolvement of social constructionism stating that "currents of social constructionism in constructivist grounded theory, has its links to social constructivism" (p. 14). Charmaz' constructivist grounded theory, as is referred to in this study, suggests that Charmaz assumes a social constructivist view underpinned by social constructionism. Charmaz (2014) considered that people subjectively construct their knowledge and reality within a social world that is influenced/informed by "social contexts, interaction, shared viewpoints, and interpretive understanding" (p. 14). Charmaz explained that "rather than explaining reality, social constructionists see multiple realities and therefore ask: What do people assume is real? How do they construct and act on their views of reality? Knowledge and theories are situated…in particular positions, perspectives, and experiences" (p. 231).

I have adopted a constructivist and social constructionist epistemological position as it acknowledges social interaction and intersubjectivity in constructing individual reality (Reich, 2009). It is compatible with constructivist grounded theory which states that the researcher is part of the research process, and the generation of data is a co-construction between researcher and participants (Charmaz, 2014). Additionally, a constructivist and social constructionist position provides the epistemological stance for understanding the broader social context in investigating human beings' experiences when interacting with AI robots and/or avatars that have been designed specifically for social human interaction.

Social Constructionism

Social constructionism has symbolic interactionist underpinnings (Berger & Luckmann, 2011), with an underscoring tenet that considers that people make their social and cultural worlds, and, at the same time, these worlds make them (Mallon, 2019). Social constructionism explores how reality is defined by human beings and society (relevant to the context and situation). Charmaz (2006) defined social constructionism as:

> a theoretical perspective that assumes that people create social reality(ies) through individual and collective actions. Rather than seeing the world as given, constructionists ask, how is it accomplished? Thus,… social constructionists' study what people at a particular time and place take as real, how they construct their views and actions, when different constructions arise, whose constructions become taken as definitive, and how that process ensues. (p. 189)

According to Holstein and Gubrium (1997), the social constructionist perspective is concerned with 'what' people construct and 'how' these social constructions develop, suggesting "considering the contingent relations between the *whats and hows* of social life" (p. 200). The social constructionist stance to grounded theory also acknowledges the 'why' questions while maintaining intricacies of social interactions. As such, grounded theory provides a method for understanding the social constructions of research participants, as well as being a method for the researcher to use throughout the research process.

Charmaz (2008) stated that while grounded theorists use a few techniques to direct generating and analysing their data, the detailed techniques unfold and emerge as a result of the research setting, and with their data, collaborators, and themselves. To what degree grounded theorists use social constructionist assumptions depends on their epistemological positioning and research paradigm. In moving grounded theory method further way from objectivism towards social constructionism, Charmaz (2008) advocated for a type of constructionism which includes "examining (1) the relativity of the

researcher's perspectives, positions, practices, and research situation, (2) the researcher's reflexivity; and (3) depictions of social constructions in the studied world" (p. 398). Furthermore, in line with the broader social constructionist perspective, Charmaz viewed actions as a core focus, developing within a socially constructed context and social structures. Additionally, the co-construction of a grounded theory, by researcher and participants, and researcher subjectivity, further identifies its incompatibility with positivist concepts of truth and knowledge as objective (Crotty, 2003).

Social constructionism acknowledges individuals as constructing meanings and knowledge as inevitably historically and socio-culturally situated and contextually bound (Kuhn, 1970; Schwandt, 2000). The social constructionist approach is concerned with the social, collective context of making meaning and socially constructed knowledge (Crotty, 1998). In other words, social constructionism focuses on the social interactions and social contexts as central to creating meaning and social knowledge being constructed and reconstructed (Ward, Hoare, et al., 2015). Social constructionism argues that reality is constructed and reconstructed both individually from the combination of experience and in connection and conversation with others (Birks & Mills, 2011; Charmaz, 2006). As highlighted by Crotty (1998), in social constructionism, the 'social' is concerned with looking to understand methods to meaning making that are utilised socially in constructing knowledge instead of seeking knowledge about social worlds.

A central differentiation between constructivism and social constructionism is the view that interpretations of meaning in reality are intersubjectively co-constructed between people who share a specific socio-cultural context and language, instead of being constructed by an individual mind (Crotty, 1998; Schwandt, 1994). Additionally, the social constructionist perspective considers the transactional and subjective nature of social interactions arise from reality and meanings. Therefore, interaction is crucial for understanding and interpreting natural and social phenomena, and using language as a way of communicating meanings dialogically (Crotty, 1998; Denzin & Lincoln, 2005; Schwandt, 1994). The concept of self can be considered a social construction, as a result of social interactions and experiences, continuously changing through new interactions and dialogues (Alvesson & Sköldberg, 2009), suggesting that human beings are shaped by and part of multi-layered cultural and contextual social world, and that meanings arise from social interactions.

Furthermore, the cultural-contextual aspects are shaped and continually evolve (individually and collectively) through the use of language as a way to describe experiences (Gergen, 1994). The epistemological positioning of the current study has focused on both social constructionism and constructivism; however, similar to Charmaz

and other grounded theorists, I explain my methodology as principally a constructivist grounded theory, in shaping my selection and use of particular grounded theory data generation and analysis methods designed to answer my research question. The chosen methodology reflects a perspective where meaning ascribed to phenomena or experiences is empirically based yet understood through the participants' interpretations. As such, relativity and subjectivity are considered a fitting and logical view of such knowledge pursuits. Social constructionism supports most of the theoretical aspects as explicated from a constructivist perspective, such as seeking to understand meanings, viewing knowledge as relative to specific socio-cultural situations, and viewing meanings as being multiple. Through co-constructions, the researcher and participants' meanings are produced for analysis, and it is through the interpretive yet specific guidelines for analysis that layers of understanding are produced of a certain phenomenon (Crotty, 1998). Constructivist epistemologies are relevant for this study, as I was interested in exploring the meanings and experiences of human beings who had interacted with an AI social robot and/or digital avatar. I recognise that these individuals had their own subjective experiences, knowledge, and realities, and that a nondeterministic approach which did not presuppose the participants' responses was needed.

Consistent with Charmaz' (2014) constructivist grounded theory, the importance of social processes and social influence are suitable and fit well with the aim of the current study; namely, to understand human beings' experiences interacting with AI robots and/or avatars, what is happening, and how it will inform the future; as well as being concerned with the reality and truth as experienced by them.

The study also takes into account the larger socio-cultural context from which people's understandings about self and other arise, in terms of their own social interactions as well as the implicit larger social and linguistic understandings that underpin these interactions. Additionally, as the researcher, I formed part of the study as I brought my personal realities, worldview, and experiences, and professional lens as a psychotherapist.

Methodology: Constructivist Grounded Theory

Mouton (1996) stated that the research design is a way to "plan, structure and execute" the research to ensure the "validity of the findings" (p. 175), which is guided by philosophical underpinnings. "A methodology refers to a theoretically informed framework for research" (Clarke et al., 2015, p. 223) and outlines how the researcher goes about practically finding out whatever they believe can be known, through data generation and analysis methods to address the research question.

At times, methodology is referred to as a framework “of logically related means and ends to guide … research design” (Giacomini, 2010, p. 129) or as a substantiative rationale for the choice in designing methods for data generation and analysis that then inform the practical procedures utilised to collect and analyse the data (Crotty, 1998). The grounded theory methodology is structured yet flexible (Tie et al., 2019) and is suitable for researching a phenomenon that not much is known about. It has the purpose of generating a theory, grounded in the data, that illuminates a process intrinsic to the research topic.

Howell (2013) noted that each methodology is connected to a philosophy which shapes the mindset and all areas of how the methodology throughout the research process is used, even to the method of coding. In selecting the best methodology and methods for this study, I considered what my relationship was between the world, the phenomenon being studied, and the topic being researched (Howell, 2013). Birks (2014) posited that “methodology is the research design that shapes the selection of, and use of, particular data generation and analysis methods to answer the research question” (p. 17). Noting that differences at the paradigm level continue into methodologies which each have definite criteria for data generation, analysis, and interpretation, Charmaz (2014) stated “which observations we make, how we make our observations, and the views that we form of them reflect these conditions as do our subsequent grounded theories. Conducting and writing research are not neutral acts” (p. 240). Constructivist grounded theory methodology was selected as little is known about the phenomenon being studied (human beings’ experiences when interacting with AI social systems therefore, I was able to construct a theory that highlights a process intrinsic in the substantive area of inquiry. This is in line with one of the main characteristics of grounded theory which aims to construct a theory grounded in the data (Chun Tie et al., 2019).

Grounded Theory Selection

Qualitative research allows the researcher to “get close to the data, to know well all the individuals involved and observe and record what they do and say” (Mintzberg, 1979, p. 586). After exploring various qualitative methodologies, I determined that this study would utilise grounded theory methodology due to the focus on the “interpretive portrayal of the studied world… [in which the] …research participants' implicit meanings, experiential views and researchers' finished grounded theories – are constructions of reality” (Charmaz, 2006, p. 10). Thus, grounded theory is concerned more in understanding rather than explaining (Charmaz, 2006). Furthermore, I selected grounded theory as it provided a means of learning about the phenomena being studied; namely, how human beings construct meanings from their actions, within the

interactional situations in which they are involved, and offered a method to construct a theory to understand the phenomena being studied (Charmaz, 2014). Grounded theory aims to generate a theory from data that were empirically derived from real-world social situations, in an iterative process of engagement with the research context and conceptual analysis (Bryant, 2009; Charmaz, 2014; Oktay, 2012). The methods of grounded theory include generating and analysing participants' subjective experiences, followed by theoretical sampling with constant comparative analysis which enables the abstraction of participants' subjective experiences into theoretical statements (Fendt & Sachs, 2008). Strauss and Corbin (1998) contended that "theorising is the act of constructing … from data an explanatory scheme that systematically integrates various concepts through statements of relationship. A theory… enables users to explain and predict events, thereby providing guides to action" (p. 25). Whereas Strauss and Corbin (1990) considered emergent theory from data as objective from the researcher, Charmaz (2014) assumed that constructivist researchers build their grounded theories by taking into account participation and interactions with both past and present people, worldviews and research approaches. Consequently, I chose grounded theory as I wanted to explore individual and collective actions, and social psychological processes of human beings' experiences when interacting with AI robots and/or avatars—focusing on what these human beings did and the meaning they made of their interactions and within their context (Thornberg & Charmaz, 2014).

Development of Grounded Theory

In considering the meanings of grounded theory and identifying a more general definition of theory in the social sciences, Charmaz and Thornberg (2012) defined a theory as the "relationships between abstract concepts and may aim for either explanation or understanding" (p. 41). Theories endeavour to answer questions by accounting for what happens, how it develops, and may focus on interpreting why it happened (Charmaz, 2014). Gubrium and Holstein (1997) suggested that qualitative researchers attend to 'why' questions "by considering the contingent relations between the 'whats' and the 'hows' of social life" (p. 200), as a form of analysis that allows for the understanding of the reasons successive actions and events happen. Charmaz (2014) then stated that "grounded theory provides both a way of analysing situated action and of moving beyond it. In contrast, most qualitative research involves 'what?' and 'how?' questions and sticks to the immediate action" (p. 228). In order to explicate the process of determining the grounded theory version that I selected for this study, it is worth briefly outlining the development of grounded theory, and the subsequent emergence of variations of grounded theory.

Grounded theory was initially developed by Barney Glaser and Anselm Strauss (1967) as a methodology for inductively generating theory that provide a valid and unbiased way of testing and refining theory by way of a full comparative analysis (Birks, 2014; Patton, 1990) with clear methods for analysing processes. Subsequently, four different methodological variations can be identified (Salvini, 2019)). First, Glaserian grounded theory (Glaser, 1978, 1998, 2005), also known as the 'classic' grounded theory approach. Second, Straussian grounded theory (Strauss, 1987), later developed in collaboration with and furthered by Corbin (Corbin & Strauss, 2008; Strauss & Corbin, 1990; 1998), required theoretical and procedural elements to have stronger consistency with the pragmatist philosophy. The third approach, constructivist grounded theory (Bryant, 2002; Charmaz, 2000, 2003, 2006, 2009, 2104; Thornberg, 2012; Thornberg & Charmaz, 2012), was interpreted and progressed by Kathy Charmaz, a student of Strauss. The fourth perspective was advanced by Adele Clarke (also a student of Strauss), who extended constructivist grounded theory and derived Strauss' (1993) ecological social world and area theory, integrating them within the cultural frames of postmodernism to generate an approach called situational analysis (Clarke, 2005; Clarke et al., 2018). Being informed by postmodernist thinking, this approach "accounts for the material environment, non-human actors, discourses, and structural elements that shape and condition the studied situation" (Charmaz, 2014, p. 220; Clarke et al., 2018). Thus, grounded theory comprises one of the most complete methodological approaches in qualitative research (Salvini, 2019). Currently, the grounded theory methodology can be viewed as a flexible framework that references Glaser and Strauss' original grounded theory, with various approaches differentiating and being formulated as a result. Glaser was informed by his positivistic, quantitative methodology training and his literature studies of line-by-line text comparison. He viewed participants' words and action as data that are objectively collected (Glaser, 1978; Glaser & Strauss, 1967; Strauss & Corbin, 1990).

Glaser (1978, 1992, 1998, 2003) focused on the development of theoretical categories that serve as "variables, assumes an indicator-concept approach, seeks context-free but modifiable theoretical statements" (Charmaz, 2006, p.127). As such, Glaser (1992) was aiming for "the achievement of parsimony and scope in explanatory power" (p. 116). Glaser highlighted the importance of utilising comparative methods and ascribes the analytic development of theory to emergence from this comparative work. However, Charmaz (2006) suggested Glaser "treats emergent categories almost as it's an automatic result" (p. 127). Glaser and Strauss (1967) first wrote about grounded theory together, in their influential book '*The Discovery of Grounded Theory'*, in which they further developed the qualitative process on methods of analysis, as well as promoting

the advancement of research rooted in qualitative data for the development of explanatory theories (Charmaz, 2014). Glaser (1992) defined grounded theory as "a general methodology of analysis linked with data collection that uses a systematically applied set of methods to generate an inductive theory about a substantive area" (p. 16). Glaser and Strauss' (1965, 1967) original work is referred to as the 'Glaserian' or classic grounded theory methodology, and provides guidelines for determining the empirical grounding of a grounded theory; that is, method and what constituted grounded theory); namely, evaluating the fit (determining if the theory fits the substantive area in which it will be used), understandability (considering if lay people involved in the area of study understand the theory), generalisability (questioning if the theory is applicable to a broad spectrum of contexts in the substantive area), and control (does the theory allow the user some control over the "structure and process of daily situations as they change through time?" (Glaser & Strauss, 1967, p. 237). However, Glaser and Strauss had different notions of the approach and rigour of data collection and data analysis, with Strauss developing a more linear approach to the research methodology (Strauss & Corbin, 1990). While classic grounded theory acknowledges the pragmatic perspective (Birks & Mills, 2015), Strauss and Corbin's perspectives (1990, 1998) recognise both pragmatist and symbolic interactionist underpinnings, with more focus on a procedural method of doing grounded theory (Birks & Mills, 2015). After Glaser and Strauss diverged in their theories, Corbin and Strauss (1990) wrote '*Basics of Qualitative Research: Grounded Theory Procedures and Techniques'* taking a prescriptive position for grounded theory, outlining the main modifications which were to the coding structure and developing their method by adding more procedures on how to collect and structure data.

These procedures make clear the differences between Glaser's and Corbin and Strauss' versions of grounded theory (Evans, 2013). However, Corbin and Strauss (1990) have been critiqued for being too rigid and inflexible, with the subsequent result potentially leading to forcing data and analysis into predetermined categories, instead of focusing on emergent theoretical categories (Charmaz, 2014; Charmaz & Thornberg, 2020; Clarke, 2005; Glaser, 1992). Corbin, more recently, has recognised that the co-construction of meaning between researcher and participant is inherent in data generation (Birks & Mills, 2015; Corbin, 2009; Corbin & Strauss, 2008, 2015); thus, adjusting her stance from an objectivist to a more constructivist view.

In summary, constructivist grounded theory, according to Charmaz (2006, 2014), considers the grounded theory research process to be fluid, interactive, and flexible; allows the research problem to guide the initial methodological adopted for data collection; and acknowledges researchers are part of what they study, not separate from it. "Grounded theory analysis shapes the conceptual content and direction of the study;

the emerging analysis may lead to adopting multiple methods of data collection and to pursuing inquiry in several sites" (Charmaz, 2006, p. 178). Additionally, continuous levels of abstraction occur through comparative analysis which is central to grounded theory analysis. Finally, researchers gain analytic guidance from how they interact and interpret comparative analysis and construct a theory. In determining my philosophical position, I reflected on the four questions posed by Birks and Mills (2015), which informed my methodological preference and, in turn, how I used the central grounded theory methods.

Regarding the nature of reality, Guba and Lincoln (2005) contended "whether or not the world has a 'real' existence outside of human experience of that world is an open question" (p. 202). In considering the different grounded theory methodological schools of thought regarding the nature of reality, I aligned with the constructivist approach, embracing a subjective and relativist position, positing that reality is understood as "relative to a specific conceptual scheme, theoretical framework, paradigm, form of life, society or culture …there is a non-reducible plurality of such conceptual schemes" (Berstein, 1983, p. 8). As ontology and epistemology are inherently linked (Birks & Mills, 2015), after determining my ontological positioning, subjectivism and relativism, I examined my philosophical beliefs about reality (pragmatism and symbolic interactionism) to guide how I could legitimately acquire knowledge about the world, epistemologically (Mills, 2014). Recognising my epistemological underpinning (constructivism and social constructionism) in this qualitative grounded theory study, my general paradigmatic orientation to theory is that of an interpretivist (Charmaz, 2014). Furthermore, a theory puts forward lines of reasoning about the world and relationships within it, can change a reader's perspective and consciousness, and can create new meanings of how the world is perceived (Charmaz, 2014). Interpretive theorising develops out of social constructionist beliefs which inform approaches such as symbolic interactionism, ethnomethodology, cultural studies, among others (Charmaz, 2014). Contemporary grounded theorist approaches continue to regard the pragmatist and symbolic interactionist theoretical stances and focus on structure/method and process as fundamental.

Constructivist Grounded Theory

The grounded theory methodology of this study is constructivist grounded theory, as outlined by Charmaz (2003, 2006). The purpose was to investigate how participants' construct, and co-construct with the researcher their experience and meaning in relation to AI social systems in their interactional and social context. Charmaz (2014) argued that "realities are multiple in the constructivist view and a multiplicity of perspectives results.

Viewers remain embedded in their worlds rather than separate from them" (p. 240); thus, recognising the broader social context of culture and historical context within individual experiences and from being in relationship and interacting with others. Acknowledging that the researcher may hold worldviews that conflict with research participants' worldviews and realities, and participants' behaviours, may expose clear differences between them and the researcher.

Charmaz (2014) argued that "relativism characterises the research endeavour rather than objective, unproblematic prescriptions and procedures. Viewing the research as constructed rather than discovered fosters researcher's reflexivity about their actions and decisions" (p. 13). Furthermore, Birks and Mills (2015) put forward that constructivist grounded theorists take on a relativist stance "where reality is relative to a conceptual framework, paradigm, form of life and is so constantly reformulated as a fluid construction" (p. 51). Constructivist grounded theorists are concerned with a relationship of reciprocity with participants (Mills et al., 2006), and recognising knowledge as contextual, provisory, and closely related to the participant and their social and physical setting (Breuer & Roth, 2003). Consequently, knowledge is considered as "constructed in processes of social interchange" (Flick, 2014, p. 78), with the research process being situated in its social, cultural, and physical context (Hanrahan, 2003) and cognisant of bias and limitations in the study (Guillaume, 2002; Ramalho, Adams, et. al., 2015). This constructivist grounded theory study explains the interactional process and experiences human beings had when interacting with an AI robot and/or avatar. It explored what was happening on an individual, subjective level as well as a wider social level, in order to understand how these social interactions can inform the future both on a micro and macro social context.

Originating from pragmatist underpinnings and shaped by symbolic interactionism, this study was open-ended and interactive, using emergent processes and being flexible so as to abductively pursue data through analytic thinking and guided by theoretical notions (Charmaz, 2014). As such, the study was consistent with Charmaz' (2014) constructive grounded theory, where the process and resultant theory of 'embodied relating' are viewed as an emergent construction of a theory. While Charmaz' constructivist grounded theory was the main methodology I followed, I reflected on other researchers that have also influenced Charmaz, such as Sally Gadow (1982), to further inform and extend my reflections; and the contemporary writings of Birks and Mills (2011, 2015) and Bryant (2017). Charmaz (2014) stated that she agreed with Gadow who assumed that "human existence essentially means embodiment and that the self is inseparable from the body… [as well as] that mind and consciousness depend upon being in a body. In turn, bodily feelings affect mind and consciousness" (p. 312).

In addition, both Gadow and Charmaz considered “bodily experience as real and as tied to self” (Charmaz, 204, p. 313). I chose to address my research questions by applying constructivist grounded theory, to explore human beings’ experiences interacting with AI social systems.

This section has focused on how constructivist grounded theory methodology contributes to the understanding of the experience of interacting with an AI social system and how it will likely inform the future in the broader relational, cultural, and socio-political contexts. Charmaz (2005, 2017a, 2017b) put forward that qualitative research approaches, such as constructivist grounded theory, allow for the opportunity to gain meaningful understanding from participants’ subjective experiences, which have not been fully explored nor understood to date. I have outlined the reason for the constructivist grounded theory approach being a suitable framework for this research, suggesting that the phenomenon being studied, in this research (i.e., the particular experiences human beings have had when interacting with AI robots and/or avatars) is new; and the meanings, actions, and processes involved need to be understood. Thus, the resultant theory provides an interpretive account of the phenomenon being studied, with the opportunity of having a wider applicability (Charmaz, 2006, 2014).

I concur with Charmaz (2014) that, like her, my symbolic interactionist notions inform my worldview, influencing what I see and how I see it, while acknowledging that these notions stay in the background until they become applicable for important analytic questions. The epistemological and methodological framework of constructivist grounded theory is consistent with the methodological stance of the symbolic interactionism, as outlined by Blumer (Aldiabat & Le Navenec, 2011; Salvini, 2019).

Summary

Through the qualitative, interpretive lens, I consider reality to be made up of people’s subjective experiences in relation to the external world and that epistemological and ontological assumptions about reality are socially constructed. As such, I sought to understand human beings’ subjective experiences within the interactional context of engaging with AI social systems. In explicating the research process, I considered the key assumptions that contributed to defining the nature of my enquiry as the ontological (subjectivism and relativism), epistemological (social constructivism and constructivism), methodological (constructivist grounded theory), and philosophical (pragmatism and symbolic interactionism) perspectives. A study that focuses on the research problems concerning human experiences interacting with AI social systems necessitated implementing a constructivist grounded theory methodology.

The nature of the research context that may make the interviews a humanising experience for the participants, and the process of analysis (particularly focused coding and the use of abductive reasoning), may help develop insights about such experiences. The purpose was to explore human being's interactional experiences with AI social systems, what is happening, and how it will inform the future.

From an epistemological stance, constructivists and many symbolic interactionists consider that it is not possible to keep the researcher separate from the participant in the generation of data. Birks and Mills (2015) stated *"*during the process of narrative interaction, researcher and participant give and take from each other. The complexity of the area of interest being explored becomes apparent and the conversation in turn gains density as meaning is clarified" (p. 55). As such, interviews become the "site of the construction of knowledge, …the researchers and the informant produce this knowledge together" (Hand, 2003, p. 17). By using meaning (as opposed to measurement) focused methodologies, such as interviewing, there is a reliance on the subjective relationship between the researcher and research participants.

The aim of constructing such knowledge, is to "reveal depth, feeling and reflexive thought", where the data are "negotiated and contextual" (Birks & Mills, 2015, p. 56). I have conducted this research, as I am troubled that there is insufficient research being conducted in the areas of human beings' experiences with AI robots and/or avatars in an interactional context and the subsequent impacts that AI social technology may have on both the people interacting with these technologies and society as a whole. I identified constructivist grounded theory as a fitting research methodology for three reasons. In the first instance, grounded theory has been connected to the field of sociological research and the study of human behaviour, experience, and the interpretation of their reality (Suddaby, 2006). Second, the main premise of grounded theory is the constructing of a theory to understand a process, rather than testing and explaining an existing theory. Thus, grounded theory is a fitting research design as there is not an existing theory that fully articulates the process as was required in this research (Creswell, 2007; Goulding, 2005; Thornberg & Dunne, 2019).

Lastly, grounded theory methodology provides guidelines for the research process. While all branches of grounded theory share common elements such as theoretical sensitivity, constant comparison, theoretical saturation, codes and memos, differences lie in the varying philosophical views (epistemology, ontology, and axiology); the rationale (deduction, induction, abduction logics) of the methods; and how the literature review is conducted (Soulard et al., 2023). The grounded theory selected for this study was shaped by the theoretical content and direction of study, which influenced the amount and type of data collection methods and resulted in various positions of inquiry while

applying rigour throughout the research process. Charmaz' (2006) statement that the aim is "to construct an interpretive rendering of the worlds we study rather than an external reporting of events and statements" (p. 184), formed the basis of this study to investigate human beings' experiences interacting with AI robots and/or avatars. Constructivist grounded theory methodology and methods are congruent with the research question as it seeks to explain a social process and generate a theory grounded in the data, with the aim of being applicable to other situations.

What do people assume is real?
How do they construct and act on their views of reality?
(Charmaz, 2014, p. 231)

Chapter 4: Research Methods

The aim of this study was to understand human being's experiences with AI robots and/or avatars in an interactional context, to determine what was happening and how it would inform the future. Over this chapter and the next I discuss the methods used to answer these questions. In the current chapter, I provide an account of the research methods and discuss how my methodological perspectives shaped the data generation and analysis process, using constructivist grounded theory procedures, as described by Charmaz (2006, 2008, 2014). I also draw on other grounded theorists, such as Birks and Mills (2011, 2015, 2019) and Bryant (2009, 2017), where applicable. I begin with addressing the ethical considerations for the study, the constructivist grounded theory-method package, my role as researcher, reflexivity, and theoretical sensitivity. Next, I outline my sampling rationale and approach, then provide an overview of the recruitment strategies, challenges, and study participants; explicating data generation in this study.

In Chapter 5, Stages of Analysis and Theory Construction, I discuss the iterative process of constant comparative analytic strategies in coding, memo writing, sampling, and integrating the analysis (Charmaz, 2008). Further, I describe how the use of grounded theory methods led to the constructed theory discussed in the results chapters.

Ethics Approval

Ethical approval was required to both protect the participants' interests, confidentiality, and privacy, and encourage me to examine my position within the research and how the study was conducted. I was granted ethics approval by the Auckland University of Technology Ethics Committee on 24th October 2019, reference number 19/390 (Appendix A). In New Zealand, ethics approval advocates for the social and cultural sensitivity of research practices which includes three Articles derived from the Treaty of Waitangi (Te Triti o Waitangi)—rangatiratanga (partnership), whai wahi (participation), and kaitiakitanga (protection).

Ethical Considerations

Rangatiratanga was implemented in the interaction between the participants and me, through encouraging mutual respect and benefit, and participant autonomy and ownership. Together, the participants and I shared our experiences and knowledge with

the aim of furthering the research in the area of social technology. The synthesis of outcomes was designed to benefit the participants and/or their social or cultural group. The information and knowledge provided by the participants are acknowledged generally (without identifying participants or their material) in the study. Whai wahi was seen in the interaction with participants, specifically during the interview wherein their principal involvement was one of sharing information (see participant information sheet - Appendix B). Participants did not have a formal role as stakeholders of the research. This research did not target participants of a particular culture or social group; neither did the study involve any participants from the vulnerable groups, as outlined by the Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants, WHO (2011).

Regarding kaitiakitanga, there were no identified risks to the participants; therefore, risk was minimal. Participants were actively protected from deceit, harm, and coercion by reminding them that they may stop the interview or recording of the interview, at any time, and obtaining informed consent (see Appendix D). Privacy was protected by confidentiality procedures including the exclusion of identifying material. Participant interviews were numbered and dated, and the participant index secured electronically, and password protected. My psychotherapy training enabled me to be aware of potential power imbalance in the interview process. I openly shared with participants about my needs, as the researcher, and my genuine interest in participants' perspectives, and the hope that the research would also benefit the wider community. Although the risk had been anticipated to be low, I prepared a Researcher Safety Protocol (see Appendix E) in anticipation for face-to-face interviews; however, all interviews were conducted via the internet on video conferencing.

In considering risks to participants, it was anticipated that the participants would experience minimal discomfort or embarrassment when discussing the research topic; although there was the potential for participants to feel vulnerable or disclose too much information regarding themselves. I have a Masters qualification and training in psychotherapy, which assisted me in being aware of each participant's level of discomfort as seen in manner, tone, and non-verbal cues which allowed me to attend to the participant's experience of being heard, respected, and not judged during the interview process. Recordings were transcribed verbatim by a transcriptionist who signed a confidentiality agreement (see Appendix G). As part of the privacy and confidentiality of the information collected, all records that had identifying material or data were either store electronically via computer with passwords or physically in a locked cabinet. Furthermore, in line with AUTEC policy, these records will be kept for 6 years after which time they will be electronically deleted or physically destroyed.

In line with the constructivist grounded theory approach, this study took into account the reciprocity between the researcher and participants, in the sharing of information (Mills et al., 2006). To that end, participants who indicated they wished to have a summary of findings will be provided with it on the completion of this study. Additionally, I will ongoingly acknowledge the study participants in any further publications.

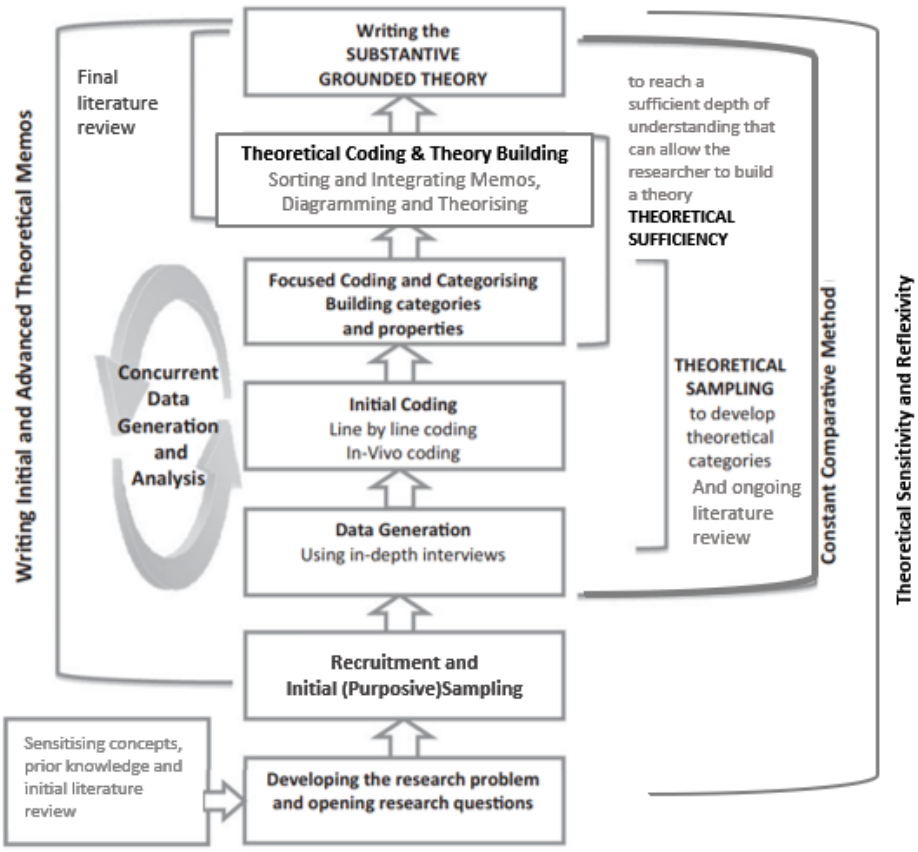
Constructivist Grounded Theory-Method Package

"Grounded theory [research] represents both a method of inquiry and a resultant product of that inquiry" (Tie et al., 2019, p. 3). Methods refer to systematic approaches and measures in which data are collected and analysed. Birks and Mills (2011) identified 'essential grounded theory methods' which begin with sampling, followed by initial coding and concurrent data generation, and the iterative process of constant comparative analysis using inductive, deductive, and abductive logic at various stages of coding, memoing, and categorisation of data. Theoretical sampling is used until theoretical saturation or theoretical sufficiency is achieved. Theoretical sensitivity, reflexivity, selecting a core category, conceptualisation, theoretical integration, and rigour further form part of the grounded theory methods package (see Figure 4.1 for a diagrammatic explanation of the constructivist grounded method employed in this study).

Grounded theory research methods are considered fluid and interactive, where the methods and processes are interconnected and are a repeat of actions and interactions that describe and inform the research process (Birks & Mills, 2019). As a researcher using the constructivist grounded theory method, I acknowledged my part in co-constructing data and meaning making in relation to my participants. The pragmatist perspective influenced my use of grounded theory methods, and through the lens of symbolic interactionism I sought to learn what the participants' experiences interacting with AI social systems meant to them; how their experiences and meanings changed; and when, why, and how they changed (Charmaz, 2014). Thus, in considering Charmaz' (2014) questions, "What do people assume is real? How do they construct and act on their views of reality?" (p. 231), I examined how humans' experiences shape and act on their perspectives of reality.

Figure 4.1

Grounded Theory Processes and Methods for This Study



Adapted from Giles, de Lacey & Muir-Cochrane (2016)

Within symbolic interactionism, the focus is on the symbols (including language and gestures) and meaning of both participants and researcher. This understanding influenced my line of questioning, interpretations, and analysis (Charmaz, 2014). Constructivist grounded theory developed as an alternative to positivism and objectivism and has a central aim of focusing on the studied phenomenon, with data and analysis considered as created from shared experiences and relationships with participants and other data sources (Bryant & Charmaz, 2007a, 2007b; Charmaz, 2014; Charmaz & Mitchell, 1996). Many constructivist grounded theorists acknowledge the rigour and flexibility in terms of including the literature review into the data analysis, and that the methodology presents a significant amount of flexibility in the construction of substantive theory (Birks & Mills, 2015; Deady, 2011). Charmaz (2014) endorsed building on the pragmatist position underlying grounded theory and in applying interpretive analysis that affirms these constructions. She explained that "constructivists study how – and sometimes why – participants construct meanings and actions in specific situations...A constructivist approach…not only theorises the interpretive work that research

participants do, but also acknowledges that the resulting theory is an interpretation" (Charmaz, 2014, p. 239). Constructivist grounded theory appealed to me as it offered flexibility through a set of methodological guidelines and tools; furthermore, Charmaz' (2006) version of grounded theory includes the researcher's worldview, thinking, and contexts, and the link between researcher and data. Through this interaction with data, the researcher draws on their own language and culture to construct meanings and shape actions. In this way, meanings are interpreted by the researcher, thereby describing the data generation in the qualitative research as interpretive in nature. Constructivist grounded theorists consider facts and values as connected, recognising the importance of being aware of their presuppositions and to address how they may impact/influence the research. Grounded theory could be considered one of the most disciplined in rigour of the qualitative approaches. Procedural precision (Tie et al., 2019), including management of data resources and an audit trail that displays the analysis and research process, upholds the quality and rigour of the study.

In the next part of the chapter, application of the constructivist grounded theory methods, as introduced thus far, will be discussed. In line with the constructivist approach, strategies of reflexive practice and theoretical sensitivity are outlined. Subsequently, I explain participant recruitment and sampling rationale and strategies, and present the participants. I address the use of multiple data sources; namely, interviewing, field notes, memos, elicited and extant data, as well as using literature as data. Finally, I explicate the data generation process and data management.

My Role as Researcher

Bryant (2009) noted that the qualitative researcher is a central element in the research landscape, "a link in the chain that reaches iteratively around data, codes, concepts and tentative theories" (p. 29). Research does not occur in a vacuum; instead, it is shaped and informed by the researcher's perspective (Charmaz, 2014; Kenny & Fourie, 2015). Consistent with the constructivist perspective, I consciously worked at developing theoretical sensitivity during the research process. Guided by specific research interests, sensitising concepts (e.g., AI, HRI, attachment theory, and social media usage), and disciplinary perspectives, I used these as tools to study the research data, coding throughout successive levels of analysis and theoretical category development from the data (Bowen, 2006; Charmaz, 2014). By acknowledging my prior knowledge, theoretical influences and analytic lenses, I recognised the need for constant reflexivity, in order to remain grounded in the data (Mills et al., 2006a). I found participants related to me as someone who was genuinely interested in learning and understanding their experiences interacting with an AI robot and/or avatar, and keen to explore and understand their

experiences co-constructively. Many participants reported not having had previous opportunities to talk with someone truly interested about their interactional experiences with these AI social systems.

Researcher Reflexivity

Various grounded theorists hold different views of reflexivity in grounded theory research (Birks & Mills, 2015). While Glaser (2001) rejected reflexivity as a technique in grounded theory methods, Strauss (1987) recognised how a researcher's reflections impact the use of grounded theory methods. More recently, Clarke (2005) stated the importance for the researcher to "become more visible and accountable for, in, and through [their] research" (p. 13). Consistent with Clarke, Charmaz (2014) argued that constructivist grounded theorists have a responsibility to integrate reflexivity as part of the method within the research design so as to identify and be aware of 'taken-for-granted' assumptions and how they shape researchers' actions and interpretations. Reflexivity in constructivist grounded theory acknowledges the researcher's subjectivity in the resultant theory (Charmaz, 1990; Ramalho et al., 2015). Through interaction with data, I drew on language and culture to create meanings and structure actions. Consequently, by interpreting these actions, the data generation and analysis could be said to be interpretive in nature. Charmaz' constructivist view recommends that researchers aim to become aware of their preconceptions and presuppositions by taking a reflexive stance and contending with how they influence the research.

In line with Charmaz' approach, I engaged in a presupposition interview at the start of the study to identify my biases and natural responses in relation to the study and the research question. Being and remaining aware of my starting assumptions and preconceived ideas through ongoing discussions with my research supervisors (Birks & Mills, 2015; Charmaz, 2014; Thornberg, 2012) meant I could grapple with their effect on my study by applying reflexivity and other procedures that enhance transparency and rigour. By using reflexivity, disproportionate bias can be mitigated by the researcher being highly aware of and continually questioning their own actions and role throughout the research process. Thus, achieving credibility with compelling and defensible results.

My presupposition interview highlighted my psychotherapeutic lens and interest in attachment theory, as well as an interest in peoples' social usage and findings from my previous research (Viljoen, 2017). As a psychotherapist, I believe my training assisted in my awareness and apprehending of my preconceived ideas both during the interviews and the analysis, throughout the research process.

The presuppositions interview was supported by writing memos to explore and reflect on my thinking and ideas, as well as discussing aspects with my supervisors. Furthermore, I actively participated in a monthly tertiary-based peer support grounded theory group of academics and postgraduate colleagues which was a valued and supportive group that provided copious amount of knowledge, reflections, and support on my research process. I was able to test my ideas and discuss my concepts and developing theory.

Conducting a constructivist grounded theory study required a mutually reciprocal relationship with participants that entailed being reflexive about any existing power disparities. It included acknowledging that as the researcher-participant relationship needs to be jointly favourable it is closely associated with the ethical principle of beneficence, which guides researchers to do no harm (Nagy et al., 2010). By adopting a more reflexive stance and proactively planning the interview time with participants, I was more aware of and able to move towards an equal position of power (Birks & Mills, 2015) through strategies such as, scheduling an interview time suitable for the participant, using a flexible and semi-structured style to empower participants over the direction of the conversation, sharing my understanding of the main factors arising, and being open to sharing details about me and my research. Proactive planning further assisted in moderating the chance that any subconscious biases or judgements may lead to negative experiences for the participants or myself.

Theoretical Sensitivity

Birks and Mills (2015) defined theoretical sensitivity as “the ability to recognise and extract from the data elements that have relevance for the emerging theory” (p. 58); and suggested it has three significant elements, namely: “1) It reflects the sum of your personal, professional and experiential history 2) It can be enhanced by various techniques, tools and strategies 3) It increases as your research progresses” (p. 58). Birks and Mills further contended that “by identifying your baseline position before you begin, you can work at consciously developing your theoretical sensitivity during the research process” (p. 59). Thus, as part of theoretical sensitivity, the presuppositions interview assisted with identifying my biases and natural responses, in relation to the study. A variety of techniques, including questioning, memoing, the flip-flop technique, and far-out comparisons (Strauss & Corbin, 1990, 1998) were used to enhance theoretical sensitivity by stimulating “reflections about the data at hand” (Corbin, 1998, p. 122) during the analysis. By enhancing my theoretical sensitivity in this study, I put forward that through my analysis and constructed theory I have not forced the data (Costain Schou & Hewison, 1998; Glaser, 1992; Pidgeon & Henwood, 1997; Sandelowski, 2000).

Reichertz (2007) stated that theoretical sensitivity is a type of abduction as the outcome is the combining of 'logic of discovery' and the 'logic of justification' within the framework of the researcher's methodological perspectives. By examining my tacit knowledge, prior experiential and discipline-relevant knowledge, as well as concepts from existing literature, my theory construction was reliant upon theoretical sensitising towards concepts distinguishable in the data (Birks & Mills, 2015; Charmaz, 2014). By immersing myself in the data, making constant comparisons, acknowledging multiple perspectives, and developing concepts, I utilised my theoretical sensitivity to transform unforeseen instances into possible theoretical developments and to abstract codes and categories that were significant for my emergent theory (Charmaz, 2014; Kelle, 2007). These abstractions advanced thinking and potential concepts that furthered data generation and subsequent interview questions, thus becoming sensitising concepts through the theory construction process (Blumer, 1969; Charmaz, 2014; Hoare et al., 2012).

Strauss and Corbin (1990) highlighted the significance of a researcher's insights with regards to what is meaningful and notable in the data. I recognised that as a psychotherapist I had the relevant theoretical knowledge from prior learning as well as prior experience. Therefore, by acknowledging this as my starting point and noting that I was not a "tabula rasa" (blank slate) (Glaser & Strauss, 1967, p. 3), I could consciously work at developing my theoretical sensitivity throughout the research process. Karen Hoare described developing theoretical sensitivity as 'dancing with data', a term inspired by the process of the researcher's path of conceptualisations, from data gathering through to constructing a grounded theory, needing to 'double back step' when moving forward in the analysis (Hoare et al., 2012); thus, developing theoretical sensitivity in the process.

Theoretical sensitivity, a central criterion in the grounded theory methods, can be enhanced by various strategies including constant comparison which involves the "weaving [of] data collection with data analysis and writing and analysing memos" (Birks & Mills, 2015, p. 59). The entirety of these strategies contributes to the enhancement of theoretical sensitivity and promotes insight and understanding in the topic areas that can form the building of a theory. The moving forwards and backwards when working with codes and categories in this study gave me the sense of 'dancing with data'. Strauss and Corbin (1990) suggested that grounded theory researchers could enhance their theoretical sensitivity through the incorporation of literature, where literature forms part of the data sources by comparing theoretical concepts with coded data; thus, potentially earning its way into the developing grounded theory (Birks & Mills, 2015). Further to the inclusion of literature, Strauss and Corbin highlighted several specific techniques for heightening theoretical sensitivity, later known as 'analytic tools' (Corbin & Strauss, 2008;

Strauss & Corbin, 1998). My focus in this study has been to use these analytic tools to enhance my theoretical sensitivity to possible theoretical constructs in the data. Furthermore, I enhanced my theoretical sensitivity to the data by using three of Corbin and Strauss' (2008) analytic tools: 1) drawing on my personal experience; 2) observing the expression of emotions and the situations from which they arose; and 3) waving the red flag (Birks & Mills, 2015). Utilising my personal experiences involved employing my tacit and professional knowledge as a conceptual comparison with the data to promote my thinking and insights, and subsequent theorising.

By including awareness of both the participants' and my own expressions of emotion, I was aware of any sensitising concepts that arose because of the inherent meanings that are linked to actions and setting. Strauss and Corbin (1990) included the technique of 'waving the red flag' for developing theoretical sensitivity, wherein the researcher acknowledges when "biases, assumptions or beliefs are [unduly influencing] into the analysis" (p. 80). In the concurrent process of data generation and analysis, I was aware of my assumptions regarding attachment theory; however, I was able to set that aside to gain greater depth and higher levels of abstraction by comparing and analysing the data, resulting in deepening my grounded theory (Birks & Mills, 2015). By applying the essential grounded theory methods and engaging in the data, I became sensitised to those factors that had meaning, significance, and merit for my developing theory. Additionally, by immersing myself in my research process and data, and using memoing to expound my assumptions and theoretical inclinations, I was able to challenge these assumptions and use the grounded theory research process as a means of developing myself and my research as part of my theoretical sensitivity. This process helped to advance my theoretical sensitivity in being aware of what I knew and becoming conscious of what I did not know (Birks & Mills, 2015). The processes of reflexivity and theoretical sensitivity assisted in advancing my levels of insight and understanding, both in terms of my philosophical and methodological perspectives, and regarding the theoretical concepts and developing theory. Theoretical sensitivity is inherently connected to my selected approach and my level of reflexivity. The value in developing my theoretical sensitivity throughout the research process has led to me having "a more integrated and abstract grounded theory" (Birks & Mills, 2015, p. 62).

Determining the Research Question

Through the process of determining my research question I selected the methods for data gathering and generation. Initially, I considered two variations in the research question; however, I took into account Charmaz's (2014) guidance to change the research questions "when you discover the other questions have greater significance in the field" (p. 16).

First, I considered a study from the perspective of those who develop AI social systems, asking: 'What is the motivator(s) for the development of AI robots and/or avatars? How do humans attach to AI?' However, this question seemed limited as it focused on one area of participants. I then considered the research question that was more focused on the human end-users of the technology and their attachment experiences: 'What does human interaction with humanised AI have on human's attachment experiences?' However, it became apparent that this question was also potentially limiting. I wanted to broaden the scope of people whose experiences I could capture to understand the interactional processes and their influence on humans' experience. Furthermore, I had become more aware of my presupposed interests in attachment theory, and it became clear that this study could potentially go beyond attachment theory to gain a deeper understanding of people's experiences. Accordingly, I revised my research question to: 'How do human beings experience AI social robots and/or digital avatars in an interactional context? What is happening and how will it inform the future?'

Sampling Rationale and Approach

In constructivist grounded theory, data are considered to be purposefully generated or co-constructed by directly interacting with participants with the aim of building a theory, rather than data being discovered (Birks & Mills, 2015; Charmaz, 2014). Having determined my research question, I then selected data generation methods and data sources, so that the data would be relevant and useful for building my theory. Data generation methods included both initial sampling and theoretical sampling, which focused on generating quality, as opposed to quantity, data to develop a constructed theory (Charmaz, 2014).

Initial Sampling

Initial sampling is where the criteria are established and how the data are going to be generated (Charmaz, 2014). In grounded theory research, initial sampling involves purposefully selecting participants who meet the inclusion criteria and can best speak to answering the research question (Birks & Mills, 2015; Charmaz, 2014; Morse, 2007). I purposefully recruited initial participants, inviting the first two participants who I knew had interacted with an AI social system, specifically, experiences interacting with an AI robot and/or avatar. My initial sampling could be said to be consistent with the pragmatist perspective, where knowledge is valued based on its usefulness and pertinence; thus, contributing to building a relevant and applicable constructed theory (Bryant, 2017). Compared to other research approaches which collect data first and then analyse them, grounded theorists generate data and conduct analysis concurrently (Charmaz, 2000, 2006; Glaser, 1978, 1998; Glaser & Strauss, 1967; Strauss & Corbin, 1990, 1998).

The analysis of the initial two interviews evoked insights, hunches, and questions which led to interviewing specific participants who could illuminate these tentative categories. I then moved on to focusing on the interchange between the data generated and analysis to inform the categories, expand their properties, and outline their implications. This process of theoretical sampling has been identified as an analytic method in qualitative research (Thornberg & Charmaz, 2014).

Theoretical Sampling

Glaser and Strauss' (1967) original definition of theoretical sampling refers to "the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyses his data and decides what data to collect next and where to find them" (p. 45). The purpose of this strategy is to be attentive to verifying and refining the building of codes and categories, while staying focused in the data generation and analysis process. The use of theoretical sampling is fundamental in grounded theory with the aim of generating further data, analysing, comparing, pursuing leads, and expanding properties and illuminating categories further to advance theory construction (Charmaz, 2014). Once I had developed initial categories and concepts, I used theoretical sampling to direct the subsequent sampling process, with the purpose of producing the highest theoretical return by using various avenues to generate data, in order to build a cohesive and theoretically sufficient theory (Charmaz, 2014). In the analysis section, I detail my use of theoretical sampling.

Participant Recruitment

I faced several challenges in my recruitment process. Through my own university I was unable to gain a form of media to place my advert, whether via a university website, newsletter, or another form of dissemination. I reached out to other universities who have engaged in research in the area of AI robots or avatars and while they were happy to discuss the topic more generally, they seemed wary of sharing too much information on their research projects. I did have one university where I had been invited to contact their research participants who met my research inclusion criteria. Unfortunately, I needed to wait for my ethics approval by which time this window of opportunity had closed. I also reached out to organisations and individuals as a means of spreading my requirement to recruit participants who had experienced interacting with AI systems. While interesting to talk with these people, no new participants were recruited during this time. I advertised for participants at the European Robotics Forum and HRI International online conference, as well as setting a specific Facebook page with the advertisement and linked to relevant groups where potential participants who met the inclusion criteria might see the advert, to no avail. Utilising my networks, I was eventually able to recruit five participants.

The recruitment of further participants was largely as a result of the snowballing technique, where a colleague forwarded my recruitment advert and participant information sheet, and that person agreed to be interviewed. Thereafter, they referred me to my next participant and so on.

Initial contact was made with potential participants by email, either by inviting them to participate and sending them the advertisement and participant information sheet or participant expression of interest, or through snowballing technique with potential participants being referred to me. Potential participants were given up to 1-month to respond. After receiving contact from a potential participant, an initial conversation was had with them about the research inclusion criteria, explaining what participation involved, and answering any questions. At this point I was able to determine whether the potential participant would be able to continue to the interview process.

Participants were selected on a 'first to respond to researcher' basis and who met the inclusion criteria; people who were likely to have had at least one experience interacting with an AI robot or avatar. Usually, potential participants were open to scheduling an interview time in this sequence of email correspondence. If this was not the case, I allowed 2-weeks for the potential participant to digest and reflect on the correspondence and the research information provided. I then re-contacted them to confirm if they were still interested in participating, had any further questions, and to schedule an interview time. There were potential participants who had initially shown interest; however, when the COVID-19 pandemic lockdowns happened they removed themselves from consideration.

Fundamental to the recruitment and interview process was respect for and protection of participants, as well as the continued ethical obligations as the researcher. Potential participants were provided with an advertisement flyer (Appendix C) and a written information sheet, allowing them to make an autonomous and informed choice to participate in the interview process voluntarily and with consent. The participant information sheet included the purpose of the research, what would happen in the research, any discomforts or risks, possible benefits, how protection and privacy would be handled, who the researcher and supervisor were in the study, and who to contact if the participant had any concerns. Participants' privacy was protected, as much as possible, by keeping their identities confidential using pseudonyms.

Using interviewing as one form of data generation, I was aware of the importance of respecting participants by establishing rapport with them. As such, I tended to explain a little about myself—my profession, my interest for this PhD, and the process of gathering and generating data, which seemed to be well-received by participants.

I was mindful of working to understand the participants' views and experiences, knowing that how I generated data would shape the content (Charmaz, 2014), in order to learn about what participants' take-for-granted, in addition to what they do not state and their actions, in order to form an interpretive understanding (Charmaz, 2014).

Informed Consent.

Once I had addressed any questions participants had and confirmed their willingness to participate and scheduled the interview, I emailed the participant consent form for them to review. I followed up with the participants the day prior to their interview to confirm their willingness to participate and remind them to sign the consent form (see Appendix D). In addition, verbal consent during the interview was obtained to return to the participant subsequently, should there be something I needed to clarify or expand on arising either from the data or theoretically.

In grounded theory research, initial sampling occurs during the planning stage. Initial sampling began by purposively recruiting five participants who met the inclusion criteria of having had an experience interacting with an AI robot and/or avatar and could answer the research question. Once, I began gathering data and working with the interplay between data and analysis, early indications and ideas guided me to further questions to ask in the next interviews, whom to ask, as well as what type of data to gather subsequently. The theoretical sampling led me to revise previous questions and add new questions to the interview protocol, as well as going back to some participants to follow-up with additional questions in emails to them and investigating extant and elicited data (Thornberg & Charmaz, 2014). Interviewing took place locally and worldwide, due to the flexibility of online video conferencing (via Zoom), with me based in Auckland, Aotearoa New Zealand.

Interviews began with pre-established semi-structured questions, and the rest of the interview was fluid and co-constructed between the participant and me. On average, each interview took approximately 60 minutes in addition to pre- and post-emails. In total, 15 participants were interviewed. As well as theoretical sampling from extant data, I sent follow-up emails to some participants to clarify points they had made in their interviews. Interview recordings from the initial five interviews were transcribed, coded, and analysed, thus guiding me in how to progress data generation. Further participants were recruited on the basis of theoretical sampling to ensure sufficient breadth and that relevant domains of inquiry were covered. The sample size of participants was always envisioned to be fewer than 30, with the resulting number of participants being dependent on theoretical sufficiency (Charmaz, 2006; Dey, 1999). I followed a similar recruitment approach as that used for initial sampling to begin theoretical sampling.

However, as the research progressed, I targeted recruitment of specific participants who could illuminate further my theoretical coding and beginning categories, and then used theoretical sampling of extant data (literature, podcasts, and a video).

My participants were a combination of male and female, from around Aotearoa New Zealand, Australia, Europe, and the UK, who reported that they had interacted with an AI social system. The central focus was on the point that participants had an interactional experience with an AI robot or avatar, whether they were the human end-user, a developer of the technology in some way, or a third-party agent using the technology as a service in their organisation.

Inclusion and Exclusion Criteria.

The inclusion criteria incorporated any person over the age of 18 years, who had interacted with an AI robot and/or avatar. Participants also needed a good command of the English language to be able to communicate effectively in an interview process. Some of the participants had English as their second or third language; however, they communicated effectively for clear understanding and meaning making when participating in the interview. All participants who expressed an interest were included, expect for one potential participant who reported they had not actually interacted with an AI robot or avatar.

Participants

As the aim is for the theory to be grounded in the data, it meant that I could not plan in advance the “how, what, where, when, who and why of data generation and collection” (Birks & Mills, 2015, p. 83). According to Charmaz (2014), a sufficient grounded theory can be constructed with minimal data sources. On the number of interviews required in a grounded theory study, Charmaz suggested that “a very small sample can produce an in-depth interview study of lasting significance. It depends on the initial and emergent research questions and how the researcher conducted the study and constructed the analysis” (p.108). Fifteen semi-structured in-depth interviews were conducted, and two participants subsequently contacted via email for further specific questions. See Table 4.1 for a summary of participant demographics.

Table 4.1

Participant Demographics

	Participants interviewed	Country	Gender	Number of social robots participants interacted with	Number of digital avatars participants interacted with	Interviewed minutes
1	Participant A	Spain	F	2	-	30
2	Participant B	Spain	F	1	-	60
3	Participant C	NZ	F	2	-	60
4	Participant D	UK	F	-	2	60
5	Participant E	UK	M	-	2	120
6	Participant F	NZ	F	2	1	60
7	Participant G	NZ	M	2	-	60
8	Participant H	NZ	M	1	-	60
9	Participant I	NZ	F	-	1	60
10	Participant J	UK	M	-	1	60
11	Participant K	FRANCE	M	1+	1+	45
12	Participant L	AUS	F	1	-	60
13	Participant M	NZ	M	-	1	75
14	Participant N	NZ	M	-	1	60
15	Participant O	NZ	M	-	2+	60

Data Generation

Birks and Mills (2015) discussed how the position the researcher takes shapes the whole research process, and it is most apparent with the researcher's relationship with the data. They also explained how the terms 'data collection' and 'data generation' can both be used in grounded theory, depending on distinct positions the researcher has in regard to how the data are acquired. For example, data generation is concerned with the researcher's direct engagement with the data sources (such as a participant) in order for data to be produced for analysis. However, when a researcher has minimal impact on the data sources, this is referred to as data collection. The value in explicating these differences is helpful when collecting and generating data in grounded theory as, the gathering of data is different. It involves the researcher either engaging with participants directly by conducting interviews, for example, or generating data personally in the form or fieldnotes or memos. "The dynamic nature of theoretical sampling in grounded theory is best supported by the active nature of these strategies. You will therefore see one or more methods of data generation used in most grounded theory studies" (Birks & Mills, 2015, p. 72). Consistent with the underpinnings of constructivism and symbolic interactionism, with regards to the generation of knowledge, I recognised that I, as researcher, and the participants that I interviewed, co-constructed data through our interactions.

Through these co-constructed interactions, I wanted to deepen the understanding of participants' experiences through a flexible process of exploring meanings and interpretations of shared experience (Mills et al., 2006; Priya, 2019). These co-constructions produced meanings for analysis; and through the interpretive, yet specific guidelines for analysis, layers of understanding are produced (Crotty, 1998). Strong grounded theory research is generated by rich data, which are "detailed, focused and full. They revealed participants' views, feelings, intentions, and action as well as the contexts and structures of their lives" (Charmaz, 2014, p. 23). In order to generate rich data, I sought 'thick descriptions' (Geertz, 1973) and adopted various data-gathering strategies such as fieldnotes, observations, elicited data, relevant extant documents, and transcribed interviews. Through these grounded theory methods, I aimed to see the research participants' views and learn about their interactional experiences. Grounded theory methods are viewed as increasing researcher flexibility when gathering data as they promote further examination of what is happening, with the benefit of coding and categorising data as they are gathered or collected (Charmaz, 2014). As such the data are modified and refined as they expand and as researcher understanding of the participants' experiences increases. In addition to the data being rich, substantive, and relevant, in creating core categories two additional criteria for data are their suitability and sufficiency for portraying empirical events. Therefore, in this study, the key was to gather and generate suitable and sufficient data, thus providing a full depiction of the topic in the research (Charmaz, 2014). In line with grounded theorists, Glaser (1998) and Stern (1994a) suggested smaller samples are not a concern, as they viewed the aim of grounded theory methods was "to develop conceptual categories and thus data collection is directed to illuminate properties of a category and relations between categories" (Charmaz, 2014, p. 33). In determining if my data were rich and sufficient, I utilised Charmaz's (2014) questions to evaluate my data:

- Have I collected enough background data about persons, processes, and settings to have ready recall and to understand and portray the full range of contexts of the study?
- Have I gained detailed descriptions of a range of participants' views and actions?
- Do the data reveal what lies beneath the surface?
- Are the data sufficient to reveal changes over time?
- Have I gained multiple views of the participants' range of actions?
- Have I gathered data that enable me to develop analytic categories?
- What kinds of comparisons can I make between data? How do these comparisons generate and inform my ideas? (p. 33)

Strategies for Generating Data

In constructing a robust grounded theory, the generating of rich data is fundamental. To ensure that sufficient data are generated, multiple data sources can be employed (Birks & Mills, 2015; Charmaz, 2014). In this study, I included the triangulation of data from multiple data sources (Denzin & Lincoln, 1998; Strauss & Corbin, 1998) to capture different dimensions of the same phenomena, as well as provide comprehensive understanding of the phenomena to construct an in-depth theory. Triangulation in this way raises the validity and reliability of the research (Denzin, 1978; Strauss & Corbin, 1998). In grounded theory, data are generated from active involvement of the researcher in the constructing of knowledge that will then be analysed using grounded theory methods. In this sense, I interacted with participants by conducting interviews and generating data, myself, by way of memos. The dynamic quality of these strategies is bolstered by the fluid nature of theoretical sampling in grounded theory. Therefore, further to generating data via interviews, fieldnotes, and memo-writing, often grounded theory researchers gather and generate data from other sources, such as literature and other materials that already exist.

Grounded theory research offers a unique position for using literature as data, “documents as data, social medial; elicited materials; and non-traditional sources” (Birks & Mills, 2015, p. 77). For this study, I employed a variety of data sources which included transcripts from interviews, fieldnotes, and memos, questions in follow-up emails, scholarly literature and novels, and podcasts published in popular media, material posted on social networking sites, and an observation from a television documentary where a human had interacted with an AI social robot, as well as extant literature and an existing video conference presentation. Over halfway through my interviewing, and as part of my theoretical sampling, I coded and analysed two podcasts from developers of AI social systems (one social robot and one digital avatar) and video documentary of a person interacting with an AI social robot, as well as a video presentation which one participant had presented previously on AI social systems and people’s experiences of them.

Additionally, I generated data from a book containing secondary data sources of research conducted between varying people and AI social systems (Turkle, 2011). In particular, the documentary (Amazon Prime, 2022) were data sources that confirmed what participants had described in their interviews, thus these data sources were considered valid and demonstrated different dimensions of the data (Mulhall, 2003) to advance the theory construction.

Interviewing

The research interview is the site for the generation of data that are socially constructed in the interchange between researcher and participant. The data are actively generated through questions and answers as a co-construction by the researcher and participant. As such, the interview process is relational and intersubjective with the aim of producing knowledge about the human situation. The research interview is conducted via the medium of language, as a tool of the interview process, with the product being in linguistic form in terms of the oral accounts and the transcribed texts that are then analysed. Through the use of narratives, participants can describe and make sense of their reality and can illuminate the meanings of the human world. Birks and Mills (2015) highlighted the importance of interviewing as a method in grounded theory research, stating it is "the principal mechanism for the generation of data" (p. 72).

While the techniques of interviewing are similar across many qualitative research studies, in grounded theory I needed to be cognisant of my interviewing technique as my aim was to generate a theory. While interviewing, I kept in mind the participants' cultural and social context and was respectful in how I asked questions and interacted with them (Charmaz, 2014). In preparing for the interviews, I completed an initial literature review which assisted me in being current with the research topic. Additionally, I became familiar with key technological terms, processes, and understandings and relevant technology, in order to be able to ask technical questions or more specialised questions, if needed, and to be able to engage efficiently with the research participant (Charmaz, 2014). By being open-ended and semi-structured in style, my interviewing reflected a symbolic interactionist stance focusing on learning and understanding the participants' views, experiences, and actions (Charmaz, 2014). When a participant expressed an experience or feeling that had potential for further exploration, I asked questions like, "That's interesting, can you tell me more about it?" or "Would you tell me how you define it?" (Charmaz, 2014, p. 69), in order to clarify or get in-depth information on the participant's intentions and meanings. There are numerous criticisms (Alvesson, 2010; Atkinson & Silverman, 2007; Briggs, 2007; Stanfield, 2011) regarding the aims and assumptions implicit in research interviewing. However, interviews are a typical form of data collection in grounded theory, and qualitative research generally (Charmaz, 2014). Additionally, interviews give participants an opportunity to reflect on their experiences and consider them anew, and to illuminate meanings and actions; thus, resulting in rich data that can inspire analytic insights for the researcher (Charmaz, 2014). Consequently, participants' meanings and views develop through reflexive progression during the interview interaction with the researcher, demonstrating co-construction in action.

Charmaz (2014) highlighted that “intensive qualitative interviewing fits grounded theory methods particularly well” (p. 85), facilitating an open-ended yet guided, in-depth exploration of the participants’ experience. Brinkmann (2015) described interview knowledge “as produced, relational, conversational, contextual, linguistic, narrative, and pragmatic” (p. 63). In other words, the “lived social and historical world of human interaction is itself something constantly produced by humans; it is also relational, conversational, contextual, linguistic, and pragmatic or action oriented” (Brinkmann, 2015, p. 63).

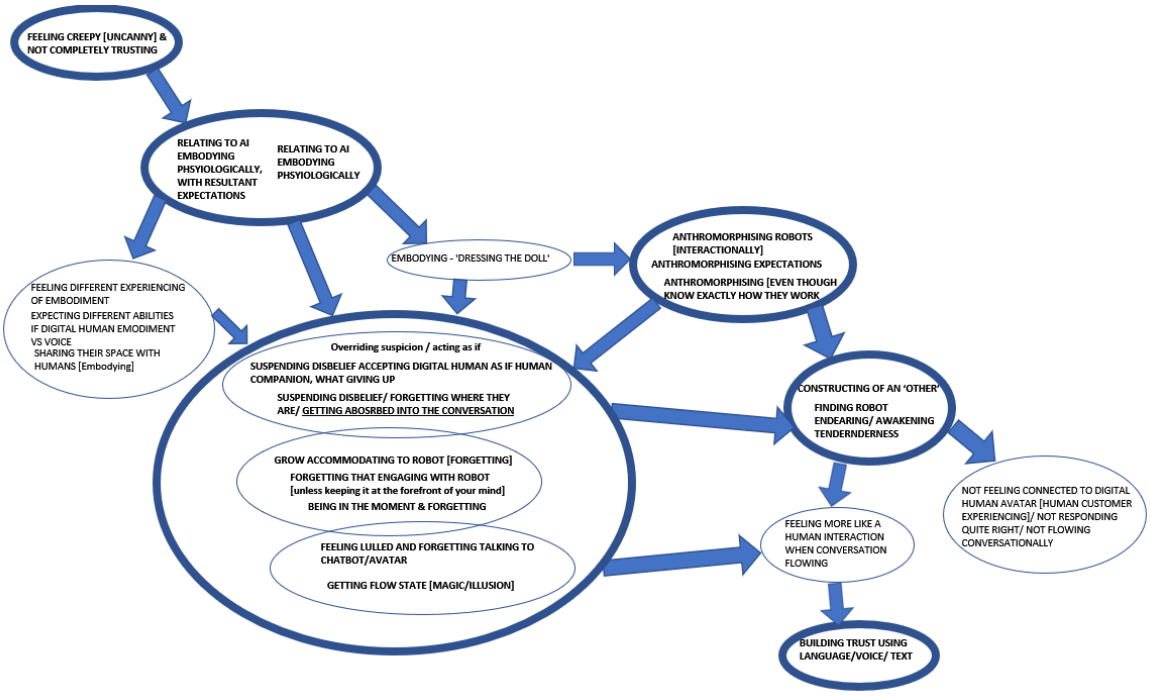
The central question of ‘What is happening here?’ (Glaser, 1978) in grounded theory, addresses the participants’ experience interacting with an AI social system. As data from early interviews were analysed, my interpretations shaped how I would ask subsequent interview questions (Charmaz, 2014). Charmaz (2014) emphasised interviewing as a central tool in generating focused data to build abstract conceptual categories, explicitly focusing on conceptual advancement and constructing a theory which distinguishes grounded theorists from other qualitative researchers who conduct intensive interviewing as their main method of data collection. Charmaz highlighted four theoretical concerns which shape the data sought and how they are gathered or collected: “theoretical plausibility, direction, centrality and adequacy” (p. 87). Key interview statements could have theoretical plausibility and through comparative analysis the researcher can determine theoretical adequacy. Theoretical centrality and direction could develop from a research idea that had theoretical plausibility or from significant coding and memo-writing (Charmaz, 2014). “When developing a grounded theory from interviews, theoretical plausibility trumps the accuracy” (Charmaz, 2014, p. 89); therefore, in grounded theory, the researcher focuses more on the theoretical plausibility of participants’ interview statements rather than on if they accurately constructed them. Grounded theory seeks to illuminate patterns and make them understandable.

Charmaz (2014) suggested that gathering data with a wide range and that are in-depth in the emerging categories advances the accuracy and theoretical plausibility of the grounded theory analysis. Grounded theorists put significance on the theoretical usefulness of interview data instead of pursuing precise accuracy. Throughout the interview process, there were times when a participant stated or described something that captured, clarified, and solidified what prior participants had signalled in earlier interviews. For example, the subcategory ‘**Growing Accommodating**’, as described by Participant L, captured earlier participants’ statements such as “*overriding suspicion*” (Participant C) and “*forgiving robots that look humanlike when they show they are 'robots' not humans*” (Participant D).

These participant statements and descriptions provided theoretical plausibility as it gave me an understanding of various other situations that I had come upon in earlier interviews (Charmaz, 2014). As I generated data and comparative analysis from my interviews, my study started to develop a theoretical direction, with the result that some participants' responses were notable whereas other statements were clustered through the process of coding and memo-writing (see Figure 4.2). The patterns that emerged informed how I conducted subsequent interviews, and my interview guide evolved with my research (Charmaz, 2014). "Clustering is a shorthand prewriting technique for getting started…[and] gives you a non-linear, visual, and flexible technique to understand and organize your material" (Charmaz, 2014, p. 184). The formations of clusters present a picture of how my codes of the studied phenomena fitted together and were linked to other phenomena. As such, clustering provided a clear visual representation in order to determine the level of significance of points within the cluster and the correlations between them (Charmaz, 2014).

Figure 4.2

Clustering of Focused Codes



As the theoretical direction developed, the theoretical centrality shaped what insights and concepts to pursue, thus focusing on the central codes and potential categories. Later interviews assisted in assessing the theoretical adequacy of my tentative categories, through conducting theoretical sampling as a means of strengthening my theoretical categories.

I would often add questions into the interview that related to my emerging theoretical concepts after a participant had initiated relevant and compelling remarks (Charmaz, 2014). Using ‘what’ and ‘how’ questions in the data gathering and collecting, I began to build an ensuing theoretical analysis. Through the iterative process of moving back and forth between the data and comparative analysis, inductive reasoning was employed. Subsequently, asking ‘when’ questions advanced the data gathering and collection toward identifying the conditions under which the human beings’ interactional experiences and processes occurred or changed. For example, I wanted to follow-up on the focused code ‘Forgetting’ and that the participant had disclosed that he forgot he was engaging with a robot. Therefore, I specifically wanted to examine the experience and process of forgetting:

> *Interviewer: What I've heard is that people know that they’re engaging with robots but sometimes if it's human looking, they forget like you?*
>
> *Participant G: Yes, unless you keep it in the forefront of your mind. I think if you get engaged in the conversation, and it depends on the ability of the robot to answer your questions but they’re pretty good now…*

By investigating the sequences and patterns, the process and their implications are determined, as well as determining meanings and actions. Charmaz (2014) noted that “from a constructivist perspective, such patterns develop as we grapple with interpreting our data” (p. 94). By listening to and analysing the audio of participant interviews, I could pay attention to participants’ nuances, language, and even tone. Charmaz contended that in attending to participants’ language, one can link the participants’ experience with the research question(s). Thus, I could determine participants’ meanings instead of making assumptions regarding their meaning. Priya (2010) suggested that the researcher’s “empathetic witnessing” (p. 479) allows participants the opportunity to share their experiences and feelings, as often they do not get the chance to talk about the interactional experiences. For example, one participant, during the interview, reconstructed the meaning of how and why they felt the way they did during the interaction with the AI social system, learning about what had been a ‘taken-for-granted’ assumption and helping them reconstruct and more deeply understand their experience. As such, emergent reconstructions include theoretically plausible data as well as shaping the theoretical direction and significance of the analysis (Charmaz, 2014).

Interviewing in Theoretical Sampling.

Theoretical sampling is considered a central principle of grounded theory research (Hood (2007). Using constant comparisons when gathering data, coding, and analysis made theoretical sampling a coherent part of the iterative process.

In line with the rationale of theoretical sampling, I was able to shape my investigation in order to develop my emerging categories. For example, most of the participants made reference to their experience of the AI robot or avatar's eyes and their humanlike appearance, referencing how they responded to them similarly to how they would respond to other humans by looking them in the eyes while interacting with them and, overall, stating that they were "*obsessing over finding spark in robot's eyes*" (Participant B), or "*recognising human shape as if human to human*" (Participant A) or "*unconsciously relating with robot as if human*" (Participant L). The theoretical sampling of this concept led to development of the theoretical category '**Unconscious Feeling**' (see Findings and Discussion chapters for more detailed discussion).

Theoretical sampling allowed for more leading questions, especially in later interviews, acknowledging the participants' responses had theoretical relevance (Charmaz, 2014). According to Charmaz (2014), focused interviewing provides numerous advantages in grounded theory studies. First, the nature of the intensive interviewing permits the pursual of new leads as well as the frequency of theoretically sensitive questions. Second, due to the iterative nature of grounded theory, conceptual categories can be constantly reviewed and modified as interviews are conducted. Thus, fostering focusing, reflexivity, and modifying and developing interview questions. Interviewing is a valued form of data generation. While I have professional experience, as a psychotherapist, to help in establishing rapport and communicating effectively with my participants when interviewing them, I heeded Birks and Mills (2015) warning not to assume that these skills will be as useful for the purpose of interviewing in grounded theory research.

Keeping in mind the emergent nature of the grounded theory research process, I needed to be flexible in how I used interviewing as a technique for generating data, considering what questions I asked and how I asked them, and how these questions may be modified or changed between and within interviews. Interviewing included being attuned and cognisant of what each participant was saying while being theoretically sensitive to what the meanings were for my developing theory and consequently adapting my interviews. I was going to pilot test my interviewing techniques in preparation for interviewing participants; however, a respected grounded theory colleague suggested there may be valuable data in that pilot interview so rather to include the interviewee as a participant thus allowing for the inclusion of the data from that interview. I am glad I did so as the interview was rich in data.

To prevent inadvertently neglecting to share important information with the participant, I had a guiding script and prompt questions (see Appendix F) that I consulted regularly. The script and prompts were, upon request, provided to participants to help them understand their rights as a person being studied.

Furthermore, it ensured that I was consciously aware of the questions I was wanting to ask and that I conducted the research in an ethical manner.

Online Video Conferencing

Birks and Mills (2015) discussed the benefits of "synchronous communication technologies such as video conferencing…to conduct interview with participants whose geographic location may make them otherwise inaccessible" (p. 73). These technologies were especially important as in Aotearoa New Zealand there were significant challenges in recruiting participants who had interacted with an AI social system and met the research inclusion criteria. Further, online video conferencing allowed me to maintain visual and audio contact with the participants in order to be able to clarify, expand on, or illuminate their words and actions (including non-verbal cues). I noticed this was apparent when participants were talking about eye contact with the AI social system and making eye contact with me at the same time. (It was especially helpful when participants wanted to show me things in reference to what they were sharing.) In order to indicate to the participants that I was engaged and interested in what they were saying, I maintained eye contact with them during the interview process, listening intently to understand their experience and views.

Showing genuine care and concern, including non-verbal communication, lets the participants know they are being heard which encourages them to share more. Throughout each interview, I sought to understand the meaning of the participants' experiences by recording and interpreting the meanings of what participants said, as well as how it was said. I observed and interpreted participants' vocalisation, facial expressions, and other bodily gestures (Brinkman, 2015). As the research progressed, the open-ended and participant-focused nature of initial interviews changed to discussing areas regarding the theoretical categories. However, when moving to theoretical sampling, I took on a more active role in the interviewing by asking more direct questions to pursue more theoretical insights while concurrently staying open to new leads. Thus, seeking focused data to build, expand, modify, and test categories and address conceptual gaps (Charmaz, 2014).

Grounded theory research promotes the researcher's continued close relationship with the data throughout the research process, unlike other types of research where data are collected as a separate task (Birks & Mills, 2015). I tended to begin each interview asking the same general question, "*How would you define and explain an AI social robot and a avatar?*" Then, as a follow-up question, I tended to ask, "*How did it come about that you could have this interactional experience with AI social robot/avatar?"*

Other general questions asked at the start of the interview allowed the participant to answer the question unencumbered by any expectations I may have had as a result of the types of questions I asked subsequently (i.e., to avoid participant bias) and included, "*Please can you tell me about your interactional experience with an AI social robot or avatar*" and "*Had you had ideas about social robots and/or avatars before you met one?"* in order to establish any preconceived view of these AI social systems. In the interview I often reiterated and used words like 'your experience' to keep them on track, within the realm of the interview topic.

Additionally, when engaging with participants, I would follow-up with more in-depth questions, in order to give them the opportunity to explore and reflect on their experiences further, "*As you look back on your experience interacting with AI social robots and avatars, are there any other events/aspects that stand out in your mind*? *Could you describe [each one] it*? *How did this event/aspect effect what happened*? *How did you respond to [the event; resulting situation]*?", "*Did you notice, if at all, what your first experience was with this AI?* [If so,] *what was it like? What did you think then*?", "*Who, if anyone, influenced you in anyway*? *Tell me about that*." If there was time, I would ask "*What is your general view of people in the world, how they are towards others and towards you*? *What was going on in your life then*?" and "*How would you describe how you view AI social robots/avatars before your experience with them*? *How, if at all, has your view of your experience changed*?" to get a sense of the social and relational context for the participant, at that time. When concluding the interviews, I asked participants, "*Is there anything you might not have thought about before that occurred to you during the interview*?" and "*Is there anything you would like to ask me*?" which allowed participants the opportunity to address any unattended aspects that they wanted to say or ask (Charmaz, 2014). Through theoretical sampling and the progression of theory construction, the semi-structured interviews moved from broader to more focused questions linked to examining more in-depth theoretical concepts and processes in the data, while staying open to new leads (Charmaz, 2014).

The interviews were transcribed for the purpose of analysis from audio recordings by the transcriptionist, and the participant's identity remained anonymous. While Glaser strongly opposed recording of interviews, constructivist grounded theorists recognise that the transcribed interviews from recordings, together with fieldnotes and memos, provide a rich data set (Birks & Mills, 2015). I personally transcribed several of the interviews and enlisted a transcriptionist to transcribe the remaining interviews, due to time constraints and for the purposes of efficiency. I checked the transcripts against the recording for accuracy.

I combined any fieldnotes with an additional strategy of re-listening to the interviews as I conducted my initial coding to ensure that I was capturing subtleties/nuances communicated in the interview. The end of the interview had a standard 'ending script' which contained important information for the participant, such as letting them know how the research would proceed and what they could expect after the interview, such as receiving a summary of the research findings.

Field Notes

Grounded theorists often undertake fieldwork as they tend to research phenomena with a sociological and human component. Fieldwork can encompass a wide variety of data generation activities, from observation and informal conversation to interviewing participants (Birks & Mills, 2015). In grounded theory studies, recording notes in the field are valuable concurrent records of actions, events, and researcher's responses to them. Birks and Mills (2015) suggested that fieldnotes should be actioned following the interviewing of each participant in order to "retain details of the physical environment, to record your immediate responses to the interaction and to capture participant non-verbal behaviour that will not be revealed through the transcription" (p. 75). All fieldnotes conducted on observations fall under the ethics approval for this study, as they relate to the interviews or have met the ethics criteria. I recorded notes intermittently while interviewing participants and made memos after the interview. I found this helpful in order to capture immediate responses and observations that may have been lost if I did not document them.

Once the interview was transcribed and I conducted my initial coding, I was able to refer to these fieldnotes for information. Charmaz (2014) emphasised how "any methods of data collection frames what you can code" (p. 136). In one interview, the participant had an activated social robot in the background which would talk during the interview and move when it seemed to think we were talking to it. In all the interviews, I made field notes of my observations of the setting and the participant, in addition to the interview, and used this as data to code, as illuminating data tends to exist in such observations (Abdi Kusow, 2003). An example of what I noticed was that I endeavoured to make a connection with the robot by waving at it, almost involuntarily.

Memos

Birks and Mills (2015) stated that memos have several purposes; however, they are particularly notable in progressing analysis as they comprise insights into the data and the analysis of that data. As the level of the analysis advances, memos are shaped by the content of previous memos, raising the conceptual level; thus, fundamentally becoming data in themselves.

It is important to be cognisant of keeping fieldnotes and memos separate in order to differentiate between raw data and a researcher's analytical abstractions (Glaser, 1978). I found memo-writing useful to help me explore notions and capture thoughts. I also became aware of how my memos evolved as the coding and analysis progressed.

Eliciting Material

Elicited material refers to data generated by participants at the invitation of the researcher (Charmaz, 2014), and where the researcher has minimal involvement with the participant and the generation of the resultant data. Such forms of data collection include questionnaires or surveys through to personal histories, logs, and diaries (Charmaz, 2014). I included elicited data by posing questions to some participants in an emailed questionnaire format, as a follow-up to the in-depth interviews (see Appendix L). I identified further points I wished to investigate leading to further data; thus, adding value to my ground theory study. Besides initiating the generation of data in terms of the questions, I did not impact the generating of these data (Charmaz, 2014). These emailed elicited materials where asynchronous because they were electronic and the participants were comfortable using this medium, which was an advantage.

Extant Documents

As opposed to elicited materials, the researcher is not involved in the construction of extant documents, which can include public records, mass media images and texts, charts and diagrams, literature, and previous qualitative material from data banks (Charmaz, 2014). Where possible, documents need to be located in their contexts. While the internet offers a multitude of opportunities for analysing documents, it also raises many methodological issues. Text can be placed in a context by employing multiple methods such as interviewing participants and using several forms of documents.

Charmaz (2014) commented that "documents that tell a story behind other texts at least suggest the social context for the analysis. Both the detail of the texts themselves and the thoroughness of the analysis figure here" (p. 53). In was important to be aware of processes that may have shaped documents which may be ambiguous, invisible, or perhaps unknowable. Awareness can be generated through asking questions of the documents, such as "How does the document reflect its author's assumptions? Which meanings are embedded in its form? In its content? How do those meanings reflect particular social, historical, and perhaps organization contexts?" (Charmaz, 2014, p. 54).

Using Literature as Data

In contemporary grounded theory, literature can be used to further enrich data analysis and is considered as a data source for theoretical sampling and a tool to stimulate theory construction (Birks & Mills, 2015; Charmaz, 2014; Thornberg & Dunne, 2019).

Corbin and Strauss (2008) considered when data are collected from literature it is classified as 'technical literature', meaning published research reports and scholarly discussions of a theoretical or philosophical quality. Birks and Mills (2015) stated that different types of documents, whether published or unpublished, can be utilised as sources of data in grounded theory research, such as "Newspapers, magazines and other media publications; Government reports; Organisational policy and procedure manuals; Personal diaries, journals, logbooks and letters; Biographies, non-fiction books and novels" (p. 79). How these materials are located will depend on the document form that is being sought. When scholarly literature is used as a data source, as part of constructing a theory, it is used in a similar way to how other data sources are used when building a theory (Glaser, 2008). For this study, documents proved very valuable as the aim of the research was to understand human beings' experiences and what is happening. Therefore, the use of literature as data was significant as theoretical sampling and constant comparative analysis with the data from interviews.

Birks and Mills (2015) suggested that information and communication technologies offer a plethora of resources which may be useful throughout the research and to be aware of the potential usefulness and limitations of these resources. The internet has become an ingrained and accepted component of most societies and there are a multitude of social networking sites, blogs, and numerous podcasts, content and video sharing sites that may contain valuable material that can be utilised in grounded theory research. I accessed podcasts, a video conferencing presentation and a television documentary, and treated these the same as other textual data; additionally, I followed ethical guidelines with regards to using this content. While these podcasts and video were in the public domain, I kept the contributors in the podcasts and video anonymous (Birks & Mills, 2015). Additionally, I coded and analysed data from a published book in which a multitude of research had been conducted on the topic of social robot and human being interactions (albeit from a different perspective). Nevertheless, it proved useful in terms of theoretical sampling for this study. Therefore, data for this study included interviews and observations, in addition to memos and fieldnotes, along with podcasts, and television documentary.

Secondary Data Sources.

Secondary data sources refer to researchers selecting data that have been collected in the past for another objective, either by themselves or by other researchers (Birks & Mills, 2015). While Glaser and Strauss (1967) had originally put forward the use of secondary data sources in grounded theory for quantitative data use, Birks and Mills (2015) suggested that data can also be generated for secondary analysis from previous qualitative studies.

There are a number of benefits to utilising data that have been collected in previous studies, such as savings in regards to time, money, and other resources; in addition, for the researcher to be able to direct their attention to the significant stages of analysis (Glaser, 2008; Goulding, 2002). Further, due to the limited resources available, in terms of potential participants to recruit, using secondary data analysis was advantageous. However, I only went to the literature for data when my theoretical sampling directed me to do so, which was especially of value in the middle to later phases of the research due to the nature of my study where there is not access to a broad range of people to recruit as participants and there has not been a substantial amount of research published. For example, Turkle (2011) had conducted many research studies interviewing a variety of people in relation to their interactions with robots, albeit from a different viewpoint, which I was able to code and analyse.

Boslaugh (2007) distinguished between primary data (that analysed by the person collecting it) and secondary data (analysed by someone else). Theoretical sampling led me to secondary data sources which proved valuable in either providing theoretical sufficiency or to fill any gaps in the analysis process. The advantage in using secondary data sets was access to a rich source of data regarding people's interactional experiences documented in literature. I analysed the secondary data sets in the same way as the primary data (from interviews), using essential grounded theory methods. Secondary data including published material of previous research on interactions between human beings' and AI robots, allowed me to make the most of researchers' earlier work which might otherwise not have been expanded upon (Glaser, 2008). Consistent with Birks and Mills' (2015) recommendation, I ensured that I applied the essential grounded theory methods in the handling of secondary data in order to reduce the possible disadvantages.

Konecki (2011) and Liebenberg et al. (2012) contended that there are possibilities for grounded theorists to use non-traditional data sources such as visual media, music, and artwork) in their research. I used a television documentary which I then translated into textual form to enable analysis. As grounded theory research emphasises the importance of data and the processes used to gather and generate data, I employed theoretical sampling strategies throughout the research as part of producing a constructed grounded theory that shows valid, logical, and analytical processes. When gathering data, I remained reflexive to my use of language and meanings as I was interpreting the data, as well as taking into account participants' worldviews and ideas both within their local and larger cultures.

Scrutinising how I gathered the data and examining that data I obtained, promoted the development of my methodological skills while concurrently locating these data. Furthermore, it assisted in the coding and categorising of data as I placed my developing analysis in its social context, which then allowed me to make detailed comparisons when coding data (Charmaz, 2014).

Summary

This chapter has outlined the constructivist grounded theory-method package used to shape the data generation and analysis process, as described by Charmaz (2006, 2008, 2014). As a researcher using the constructivist grounded theory method, I have detailed my role in co-constructing data and meaning making in relation to my participants.

Ethical considerations were addressed, as well as my use of reflexivity and theoretical sensitivity. I provided an overview of the recruitment strategies, challenges, and study participants, and explained the process of data generation. In the next chapter, I describe the stages of analysis and theory construction for this study.

While actors act towards their environment, they are both socially and materially a part of their surroundings… subject, cognition, knowledge and theory all need to be conceptualised as a dynamic process resulting from activity relating to actors with environment.
(Strübing, 2007, p. 595)

Chapter 5: Stages of Analysis and Theory Construction

Introduction

This chapter outlines the stages of analysis in the essential grounded theory method; explicates comparative analysis, including inductive, deductive, and abductive reasoning; and considers grounded theory analysis in general. Theoretical sampling and coding are described as employed in this research. I outline my used of software, memo-writing, diagramming, and storyline to assist in the management of data and analysis. I then explain my position for advanced analysis and my perspective on theoretical sufficiency and study rigour. I conclude with an evaluation of the grounded theory method.

In line with grounded theory methodology, I commenced qualitative analysis with an openness towards and curiosity about the data. I engaged a simultaneous process of analysis, data generation, and memo writing to document analytic thinking and methodological assumptions in relation to developing codes and theory. The data analysed were generated from the transcribed interviews and memo-writing. Codes were developed and categorised from the systematic analysis of early data through initial sampling (Charmaz, 2006). Through the process of constant comparative analysis, codes were grouped and refined into categories with descriptive accounts of the processes and interactions between categories (Corbin & Strauss, 2008). Categories were, in turn, reviewed and refined, to identify a core category leading to theory construction. Conceptualisations were constructed using the constant comparative analysis to describe processes and actions. Through the process of revealing action and conditions shifting action, a succinct conceptual theory was developed with the intention that the 'process' described may be useful to other contexts. The final phase of coding provided descriptive accounts of the processes and interactions between the categories, subcategories, and properties (Corbin & Strauss, 2008), with a comprehensive and explanatory presentation of the findings discussed in Chapters 6 to 9.

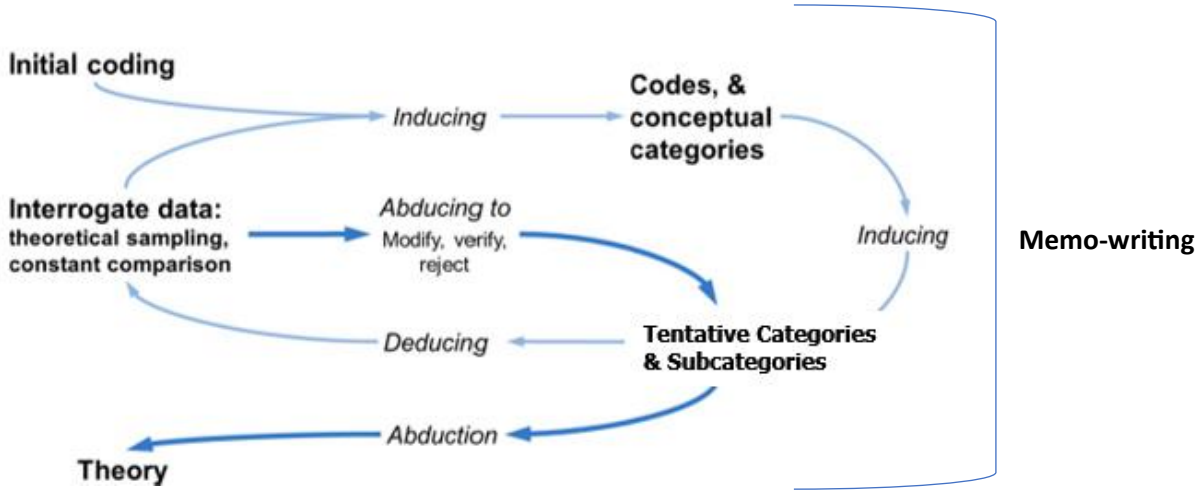
Constant Comparative Analysis

Central to the grounded theory method is the use of constant comparative analysis techniques and the researcher's engagement (Charmaz, 2006), as both make up the core of the method. The key to constant comparative analysis is the sequential levels of

conceptualisation; thus, being able to identify consistencies and differences with the purpose of repetitively fine-tuning concepts and theoretically relevant categories. The repetitive process included inductive, deductive, and abductive reasoning (see Figure 5.1).

Figure 5.1

Reasoning in Grounded Theory Research



Adapted from Ward (2016)

Through making constant comparisons and interacting and interpreting the data, codes, and categories, the researcher's theoretical understanding develops as the analytic features of the research categories are defined and subsequently rigorously analysed. My analysis became distinctly more theoretical when I applied the question "of what theoretical category are these data an instance?" (Charmaz, 2006, p. 179), and I probed the relationships between the categories, as well as key factors of being human, of human social relations, and from individual to broader social actions and structures. Throughout the coding process, I used the constant comparative technique where I compared data with data, data with code, and code with code, to identify similarities and differences (Glaser & Strauss, 1967).

Employing the constant comparative techniques within the initial coding process, I sorted and clustered the initial codes which led to identifying codes that needed to be revised, or developing new codes that were more applicable/pertinent, or combining identical or similar initial codes (Thornberg & Charmaz, 2014). I resonated with Charmaz' (2006, 2014) view that 'interaction is interpretive': as human beings we understand our experiences, assess what is happening, and use language and culture to make meanings and put actions into context. Throughout the research process, the grounded theory methods capture a researcher's passing thoughts and instant questions; and, from making provisional interpretations in initial coding to memoing to building a theory,

the researcher is prompted to advance interpretative notions/concepts into substantiative and analytical writing. Charmaz' (2006) focus on constructivism "loosens grounded theory from its objectivist foundations" (p. 180). While researchers have raised concerns regarding the reliability of qualitative data of an interpretive quality used for data collection, Charmaz posited that grounded theory holds an interactive value and by employing reflexivity, in more recent grounded theory versions, the researcher remains close to the data, interacting with it and developing concepts/notions to promote abstract interpretations. While I have endeavoured to present the participants' views and experiences, there is acceptance that "we shape the data collection and redirect our analysis as new issues emerge" (Charmaz, 2003, p. 271).

Mead's concept of emergence recognised that present and past reality differ, and that new influences or factors in the present experience lead to new interpretations and actions (Charmaz, 2014). Indeed, Mead (1938) concluded that "e*mergence*, relativity, and pragmatism are three conceptions which belong to this modern world" (*p.* 642). Pragmatist concepts such as emergence (Mead, 1938) and abduction (Peirce, 1931-1958) are acknowledged in this study, as I recognised that unexpected concepts may result from the data and analysis that could not have been predicted in advance. Understanding that these indeterminate concepts would influence and direct further iterative data generation and analysis, theoretical sampling supported following new leads, testing ideas, and filling gaps in the data. Thus, utilising concepts about emergence facilitated useful, but unexpected, directions of grounded theory research in order to understand the data and construct theory (Charmaz, 2014; Kenny & Fourie, 2015).

While there are differing opinions on whether grounded theory reasoning is purely inductive (Glaser & Strauss, 1967), inductive and deductive (Corbin & Strauss, 2008), solely abductive, or some varied combination (Strübing, 2007; Ward et al., 2016), Charmaz (2008) posited that in constructivist grounded theory "abduction allows for intuitive interpretations of empirical observations and creative ideas that might account for them" (p. 157) in the iterative process of analysis. What differentiates abduction from other types of reasoning, is the generating of a tentative theory to explicate observed data, using creativity and the development of new thinking/insights to extend beyond inductive reasoning (Morgan, 2020). Abductive reasoning accounts for new analytic insights, by elevating the level of abstraction of the analysis, thus increasing its "theoretical reach" (Charmaz, 2008, p. 168), with "abduction seek[ing] a theory" (Tavory & Timmermans, 2019, p. 537). Constant comparative analysis in grounded theory is an iterative and circular process that uses three forms of reasoning for analysis and building a theory: induction, deduction, and abduction (Figure 5.1).

Inductive reasoning involves data generation and code construction; deductive reasoning includes testing, refining, and rejecting ideas and concepts, and uses theoretical sampling to gather further data. Abduction entails raising data to a conceptual level. Utilising constant comparative analysis was a useful technique for incorporating literature with my memos, data, codes, and categories in the theory building research process (Ramalho et al., 2015).

Induction.

When constantly comparing, via analysis and decision-making, I used inductive and abductive reasoning. Bryant and Charmaz (2007) defined inductive reasoning as "a type of reasoning that begins with study of a range of individual cases and extrapolates patterns from them to form a conceptual category" (p. 608). Induction includes the generation of data and the construction of codes. Inductive reasoning occurs when a pattern is identified from a sequence of empirical and individual events from which a general statement can be determined that is plausible and tentative (Thornberg & Charmaz, 2014). In other words, inductive reason is "the process of observing a number of instances in order to say something general about the given class of instances" (Brinkman, 2015, p. 224).

Deduction.

Deduction includes advancing data to a conceptual level to scrutinise, modify, and possibly reject provisional concepts; in addition to establishing whether further data need to be gathered to develop and analyse concepts for theoretical sampling. Deduction is concerned with predicting what will occur or result in a given situation by applying a broad statement or rule (Thornberg & Charmaz, 2014). Brinkman (2015) stated that both inductive and deductive reasoning assume that researchers "have some stable entity we can analyse repeatedly in a number of cases to build general knowledge (induction) or that we already have general ideas from which we can deduce particular consequences to test (deduction)" (p. 225). However, in referring to "the unpredictable conversational world of human beings" (Brinkman, 2015, p. 225), utilising abductive reasoning as a dynamic approach is more useful.

Abduction.

Julius Pacius first used the term 'abduction' in the late 16th century from Aristotle's translation of the term 'apagoge' meaning the "third form of inference after induction and deduction" (Bryant, 2009, p. 29). Peirce (1960; 1979), a pragmatist philosopher, further developed the concept of abduction, along with others (Anderson, 1987; Reichertz, 2007; Schurz, 2008; Walton, 2004).

Charmaz (2006) defined abduction as:

> a type of reasoning that begins by examining data and after scrutiny of these data, entertains all possible explanations for the observed data, and then forms hypotheses to confirm or disconfirm until the researcher arrives at the most plausible interpretation of the observed data. (p. 186)

Abductive reasoning involves selecting or constructing a tentative theory as a useful contender for further inquiry that explains a particular empirical case or set of data better than any other potential theory (Thornberg & Charmaz, 2014). The outcomes of abduction are always tentative, open to being modified in light of new data and to better premises or explanations. Abduction entails the researcher continually moving back and forth between data and established categories; as well as developing, knowledge, or hypotheses, then comparing and interpreting in order to identify patterns and optimal explanations (Thornberg & Charmaz, 2014). Constructivist grounded theorists acknowledge and utilise the analytical strength of the constant interchange between induction and abduction throughout the research process, utilising knowing and including literature as a means of achieving many potential lenses (Thornberg & Charmaz, 2014). Thornberg and Charmaz (2014) referenced the value of acknowledging established theories and research findings to build on knowledge but engaging with pre-existing forms of relevant literature to have greater insight. To make good abductive reasonings in my research, I needed to take into account my prior knowledge, and be flexible and open to other beliefs and thinking (Kelle, 1995; Thornberg & Charmaz, 2014).

The concurrent generating of data and analysis of codes and categories is an essential method of grounded theory, distinguishing it from other qualitative research designs (Birks & Mills, 2015). Theoretical sampling and the ongoing gathering and generating of data occur as a result of concurrent data generation and the distinct notional stages of data analysis. Through the iterative analytical process of constantly comparing while gathering and generating data, I developed and advanced categories that were rich with meaning, inclusive of properties, and explicated the dimensional variance of categories (Birks & Mills, 2015). For example, I considered the subcategories ‘Projecting’ and ‘Anthropomorphising’ and whether they overlapped. However, as I refined the categories and subcategories, I determined that these were distinct subcategories with their own properties.

After investigating the reasoning of grounded theory, Reichertz (2007) suggested that grounded theorists employ a combination of inductive and abductive reasoning, which explains the ‘conceptual leaps’ of analysis that result as the grounded theory becomes more conceptually abstract.

Theoretical comparisons can give insights for theoretical sampling to determine any variation, and theoretical comparisons occur between categories and other abstract concepts in order to illuminate potential properties and dimensions that may have not been evident before. Another theoretical comparative technique is systematic comparison, where an incident in the data is compared to either a remembered experience or the literature, thus attending to concept comparison rather than individual comparisons. By using systematic comparisons, the researcher can sensitise to potential properties and dimensions in the data that may not have been identified previously. The focus is on how often the concept arises and what its properties are like in changing conditions. While the actual properties arise out of the data, this comparative technique assisted me to recognise the properties that I might have otherwise missed. As such, I made theoretical comparisons founded on my experiential knowledge or from the literature; for example, examining how humans interact with each other and the processes and actions that follow (Fonagy, 2001; Holmes; 2001).

Grounded Theory Analysis

As well as being iterative and comparative in method, grounded theory is an interactive process. Charmaz (2014) explained how the researcher moves into an interactive space, engaging more deeply with the data by acting on the data which results in the researcher's continued engagement with the data. This process leads the researcher to developing new analytic questions while grappling with learning about the studied phenomenon. The interactive space allowed me to scrutinise my earlier preconceptions and hunches and I began to make analytic sense of fragments of data, with tacit meanings becoming evident and determining new links between the data (Charmaz, 2014).

The first phase of coding, initial coding, directs the researcher to study their data intently, line-by-line, and to start conceptualising their notions. Initial coding guides the direction of analysis in the early stages of grounded theory research while remaining open to the data, giving preliminary insights for exploration and analytical examination in order to identify new insights and subtle meanings. The second phase, focused coding, allows for the sorting and combining of large quantities of data and attending to initial codes that suggest analytic significance. Theoretical coding is "a form of coding to integrate and solidify the analysis in a theoretical structure" (Charmaz, 2014, p. 19). Therefore, coding guides how researchers view their data and assists with gaining more in-depth understanding of the studied phenomenon. Coding connects data gathering with building an emergent theory, providing an analytic framework for the grounded theory study. Employing grounded theory coding, I interacted with the data recursively, leading me to study the data by coding and learning about participants' implicit meanings.

The subjectivity of coding allowed me to ask new analytic questions and utilising constant comparative analysis codes were modified and advanced as my analytic understandings developed (Charmaz, 2014). Particular codes solidify meanings and actions in the data; with memo-writing advancing the development of concepts, insights, and analysis by comparing data, exploring notions about codes, and guiding further data-gathering. As analysis and coding progresses, particular codes are raised to potential conceptual categories (Charmaz, 2014).

Data Analysis

My methodological perspectives influenced how I analysed the data and constructed the theory. As discussed earlier, the essential grounded theory methods are iterative; therefore, through analysis of early data from initial sampling, initial codes were developed (Charmaz, 2006, 2014). Later, theoretical sampling was employed to build on the earlier conceptual analysis and continued until theoretical sufficiency (Dey, 1999). Through constant comparative analysis, the conceptualisation of processes and actions were explained. Birks and Mills (2015) offered two 'rules' in grounded theory analysis; first "that everything is a concept" (p. 86), which can be understood as "descriptive or explanatory idea, its meaning embedded in a word, label or symbol" (Holloway, 2008, p. 43). Concepts in grounded theory are connected in terms of their purpose within the analytical process and various levels of complexity. Second, data analysis progresses in terms of the research questions, aims, and analysis, keeping the substantive topic of enquiry in mind in order to stay attentive and advance analytical depth and integration. I was aware that the inductive nature of the grounded theory process could result in me being overwhelmed by the amount of data; thus, I found it useful to constantly forefront the research question by putting it at the top of my computer monitor. Returning to the research question and checking if what I was analysing worked towards answering the research question helped to keep me focused and identify when I was getting side-tracked.

Theoretical Sampling

Theoretical sampling is a grounded theory strategy used to gain selective data in order to refine and develop core categories. Birks and Mills (2015) defined theoretical sampling as "the process of identifying and pursuing clues that arise during analysis in a grounded theory study" (p. 68). Theoretical sampling commenced out of the initial coding and analysis and additional data were collected to contribute to generating a theory. Theoretical sampling allowed for leads to be identified and followed on from the analysis and progress to sampling new participants or material that provided relevant information to develop a theoretical category.

Thus, theoretical sampling is a central technique accountable for the “making the process emergent” (Birks & Mills, 2015, p. 67) and for raising theoretical sensitivity, resulting in being “attuned to… clues and their significance for your developing theory” (Birks & Mills, 2015, p. 68).

This type of sampling is “theory-directed sampling” (Strauss, 1987, p. 276) which provides different ways of examining concepts and their meanings in the building of a theory. Throughout the research process, theoretical sampling offered me direction as what data to gather next, in conjunction with constant comparative analysis that carried on until the categories had been fully developed and reached a level of sufficiency (Birks & Mills, 2015). Theoretical sampling also enabled me to verify, clarify, and expand developing categories (Charmaz, 2014). Furthermore, theoretical sampling identified gaps in and between categories (Corbin & Strauss, 2008). Theoretical sampling was used throughout the research process to advance abstraction and theoretical reach, highlighting further theoretical connections, subsequent to analysing and coding more data and filling out properties (Charmaz, 2014).

Theoretical sampling assisted me in checking, clarifying, and explicating the boundaries of my categories and outlining links between categories. The benefit of using theoretical sampling was that it included analysing documents, conducting observations, as well as interviewing or reinterviewing or eliciting data in follow-up questions following an interview (Charmaz, 2014). In line with the rationale of grounded theory, theoretical sampling is emergent, with ideas and insights evolving and shaping what a researcher investigates and the type of questions asked during the process of theoretical sampling. Furthermore, this technique assisted in the process of analysing implicit views and meanings. In grounded theory, theoretical sampling advances analysis and keeps the researcher moving towards building theoretical categories. As such, theoretical sampling can identify analytic problems. I recognised times in my research process where I felt uncertainty and ambiguity, and I was reminded numerous times by supervisors and the grounded theory group to ‘trust the process’. Theoretical sampling advanced my analysis by promoting the pursual of analytic leads, resulting in identifying appropriate properties of my categories, fine-tuning my categories, and explicating the analytic connections between and among the categories, providing more in-depth and abstract analysis that is more generalisable. It also grounded my assumptions in the data and refined my theoretical statements (Charmaz, 2014). As I focused on theoretical categories, theoretical sampling directed me to sample across substantive areas and fostered the raising of my theory to a formal, more abstract level spanning various substantive areas (Charmaz, 2014).

Charmaz (2014) noted that “theoretical sampling ensures that you construct full and robust categories and aids you in clarifying relationships between categories” (p. 200). As a result, theoretical sampling helped me to make distinctions about participants’ experiences and form new questions to advance my analysis. By employing theoretical sampling, I used abductive reasoning to examine all potential theoretical explanations for the observed data. Subsequently, abstract conceptualisations were tested for each explanation until the most plausible theoretical interpretation of the data was determined (Charmaz, 2014). Using abduction as part of the grounded theory method, potential hypotheses can guide researchers “in the most useful, emergent and often unanticipated theoretical direction to understand their data” (Charmaz, 2014, p. 201).

Coding

Charmaz (2014) described coding as “naming segments of data with a label that simultaneously categorises, summarises, and accounts for each piece of data” (p. 111). Grounded theory promotes examining action and processes, with coding providing the tools for fragmenting, sorting, defining, and synthesising data from interviews, documents, and other texts. Grounded theory coding moves beyond the ‘concrete’ data to a more interpretive depiction. In this study, the codes stuck closely to the data, demonstrating actions, and showed the sequence of events from the participants’ experience and perspective. In line with grounded theory focusing on emergence, I created codes from the data as opposed to coming from presuppositions (Charmaz, 2014).

Charmaz found grounded theory especially useful for the in-depth and detailed analysis of data by fragmenting them into parts. Grounded theory coding allowed me to begin conceptualising what was happening in the data. Charmaz promoted the coding of full interview transcriptions as it advances the researcher’s notions and gives a deeper understanding which could have otherwise been missed or forgotten. Additional advantages of transcribing whole interviews and field notes include being aware that the first coding of data does not need to be the last one, where rich data can generate a multitude of questions and analyses; iteratively returning and recoding a set of initial data; and igniting new notions. Irrespective of the type of data source that has been generated or collected, the analytical process stays the same as data are initially analysed for incidents (i.e., a broad term referring to recurring actions, characteristics, experiences, phrases, explanations, images and/or sounds). Within incidents, I detected the concepts that could then be coded and used to determine conceptual reoccurrences and similarities in the data (Birks & Mills, 2015). In grounded theory, frequently a code is heightened into a resultant name of a category due to it having an ‘inherent conceptual grab’ (Glaser & Strauss, 1967) for the researcher.

Throughout the grounded theory study there is an interplay between the gathering of data and coding. I used coding to investigate and engage with the data, in addition to asking analytical questions of the data to identify what was happening in the data. Thus, I began to contend with what it meant, via the process of coding (Charmaz, 2014). Codes are formed out of the language of the data, and usually take the form of gerunds and/or in vivo codes. 'In vivo' is a Latin term that means 'within the living'; as such, in vivo codes refer to a participant's words that have been utilised to represent a wider concept in the data. Grounded theorists consider in vivo codes as codes that are verbatim special terms that participants have used and are "symbolic markers of participants' speech and meanings" (Charmaz, 2014, p. 134). It allows these codes to reveal participants' meanings and discern their emergent actions, as well as making comparisons with data and developing categories.

Gerunds, verbs that are employed as nouns and always end with 'ing', are used for coding for the purpose of ascertaining the process in the data in the initial coding phase. In addition, gerunds are used to concentrate on the "participant's experiences as a source of conceptual analysis" (Birks & Mills, 2015, p. 90). Constructivist grounded theorists (Charmaz, 2006) view coding as consisting of no fewer than two phases: initial coding and focused coding. That said, coding is not a linear process, as researchers move back and forth between the various phases so as to theoretically sensitise. Usually, initial coding is conducted more at the beginning, as opposed to the end, of the study (Thornberg & Charmaz, 2014). Initially, I compared statements and incidents within the same interview and then compared statements and incidents in other interviews (see Appendix I). I also compared elicited and extant data with data from earlier interviews. Analytical methods describe the various stages of analysis using different language. The different stages of these analytical processes correlate with the researcher's level of conceptual abstraction being developed at the time. Charmaz (2014) referred to initial, focused, and theoretical coding which are directly connected with levels of conceptual development (which Birks and Mills (2015) aligned with low-, medium-, and high-level conceptual analysis).

Birks and Mills (2015) described initial coding as the processes that propels grounded theory analysis by developing low-level conceptual analysis. As coding progresses to focused coding (Charmaz, 2014), where initial codes that show analytic strength are elevated to provisional categories to develop, the level of conceptual analysis is raised (Birks & Mills, 2015). Theoretical coding refers to "codes researchers draw on from prior theories or analytic schemes and use to integrate the categories of their analyses" (Charmaz, 2014, p. 345).

Birks and Mills viewed a completely integrated grounded theory as being a high-level conceptual construction that has explanatory strength by demonstrating 'advanced analytical processes'. Due to the iterative and circular nature of grounded theory, I alternated between phases of coding throughout the research process.

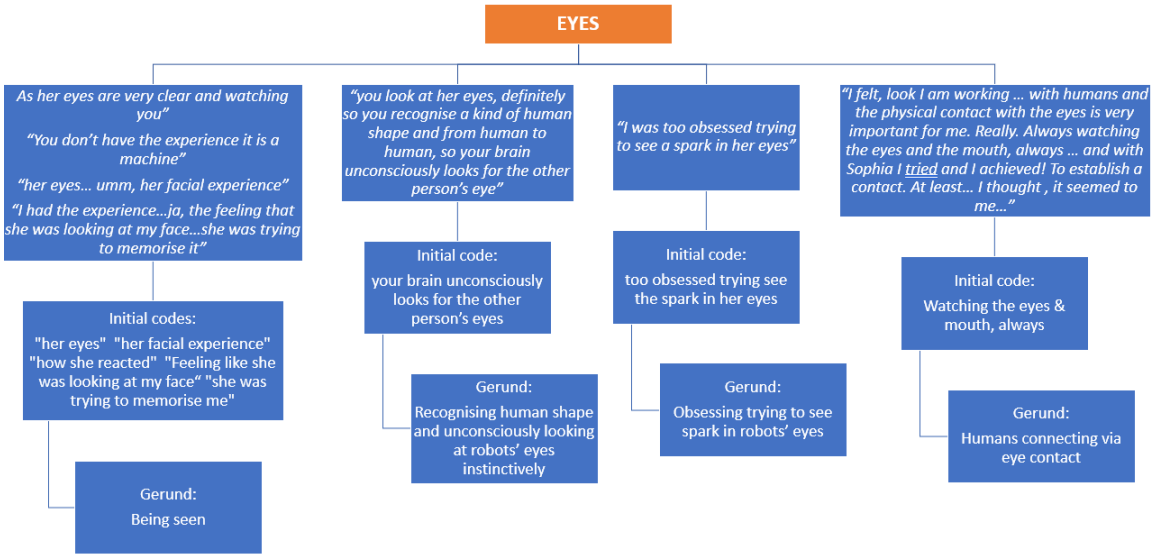
Initial Coding.

Initial coding was undertaken analysing transcriptions, fieldnotes, and memos, and formed the first part of analysis where data were broken-up and codes inductively generated by identifying and labelling key words or phrases that highlighted social and psychological processes in the data. Through the process of initial coding, I began to categorise and assign meaning to look for developing patterns which informed the direction for further data collection/generation. When conducting initial coding I stayed close to the data, while comparing and being open to interpreting processes and actions in the data.

I conducted line-by-line coding for my initial coding, using gerunds throughout, which assisted me in understanding participants' implicit meanings and actions, motivating constant comparison with the data, and providing potential leads and developing processes in the data for pursual and further investigations. For example, the in vivo codes in one incident led to my initial code being condensed into "*Watching the eyes and the mouth, always*" (Participant A) and the gerund 'Humans connecting via eyes' (see Figure 5.2).

Figure 2.2

Example of In Vivo Codes, Initial Coding, and Use of Gerunds from interview transcriptions (blue) and Subsequent Focused Code (orange)



Subsequently, I sought for and coded the condensed meanings and actions of each incident where participants had mentioned the robot or avatar's eyes and what these terms involved (Charmaz, 2014). Charmaz (2006) noted that grounded theorists ask analytical questions while coding data as a way of having flexibility in analysing data:

> 'What is this data a study of?', 'What category does this incident indicate?', 'What is actually happening in the data?' (Glaser, 1978: 57), 'What is the participant's main concern?' (Glaser, 1998: 140), 'What do the actions and statements in the data take for granted?', 'What process(es) is at issue here? How can I define it?', 'How does this process develop?', 'How does the research participant(s) act and profess to think and feel while involved in this process?', 'What might his or her observed behaviour indicate?', 'When, why, and how does the process change and what are its consequences?' (p. 51)

Charmaz (2014) described initial coding as "the early process of engaging with and defining data. Initial coding forms the link between collecting data and developing an emergent theory to understand and account for these data" (p. 343). Glaser and Strauss (1967) discussed initial coding as a process being used to 'fracture the data', where comparisons are made incident with incident, identifying seeming phenomena and beginning patterns, and starting to make comparisons between the codes used.

Throughout the initial coding I stayed open to investigating any theoretical possibilities identified in the data; and through the process of initial coding and comparing and analysing data, I began to consider my core conceptual categories. Thus, the openness of initial coding ignited my thinking and enabled the emergence of new ideas. While earlier grounded theorists posited using initial coding without any preconceived notions (Glaser, 1978, 1992), Charmaz acknowledged the value in open-mindedness and emphasised that researchers have prior knowledge and experiences. Therefore, I worked to stay open to seeing what I coded and where it led me, using reflexivity to consider how my past shaped my worldview and how I saw the data. As such, sensitising concepts helped me initiate coding my data and provided starting points for analysis. According to Charmaz (2014) "sensitising concepts from symbolic interactionism include action, meaning, process, agency, situation, identity, and self" (p. 117).

In the first two interviews, I coded line-by-line and identified key meanings which I then coded in gerunds, in order to avoid static labels as initial codes. Charmaz (2014) put forward that by coding actions researchers are less likely to make conceptual leaps and implement extant theories prior to the required analysis. Codes are tentative and can be reworded to advance their fit with the data so that they are compelling, capturing, and encapsulate meanings and action to a better extent.

While conducting initial coding I stayed open and close to the data, worked to keep codes short and precise, used gerunds to keep the actions, compared data with data, and endeavoured to move quickly through the data (Charmaz, 2014). The benefit of using grounded theory strategies meant that by conducting initial coding I could identify gaps in the data from the beginning stages of the study to inform the analytic process.

Charmaz (2014) identified coding strategies that, when used flexibly, can result in coding leading to building theoretical categories through step-by-step analysis. These strategies include "breaking the data up into their component parts…; defining the actions of which they rest; looking for tacit assumptions; explicating implicit actions and meanings; crystallizing the significance of the points; comparing data with data; identifying gaps in the data" (Charmaz, 2014, p. 125). Line-by-line coding provided me with insights into what types of data to gather next. Therefore, where required, I returned to previous participants to clarify aspects of data and determine new participants or other data sources that could illuminate the experiential process (Charmaz, 2014). I did this with two participants (Participant J and Participant L). For example: One of my follow-up questions asked:

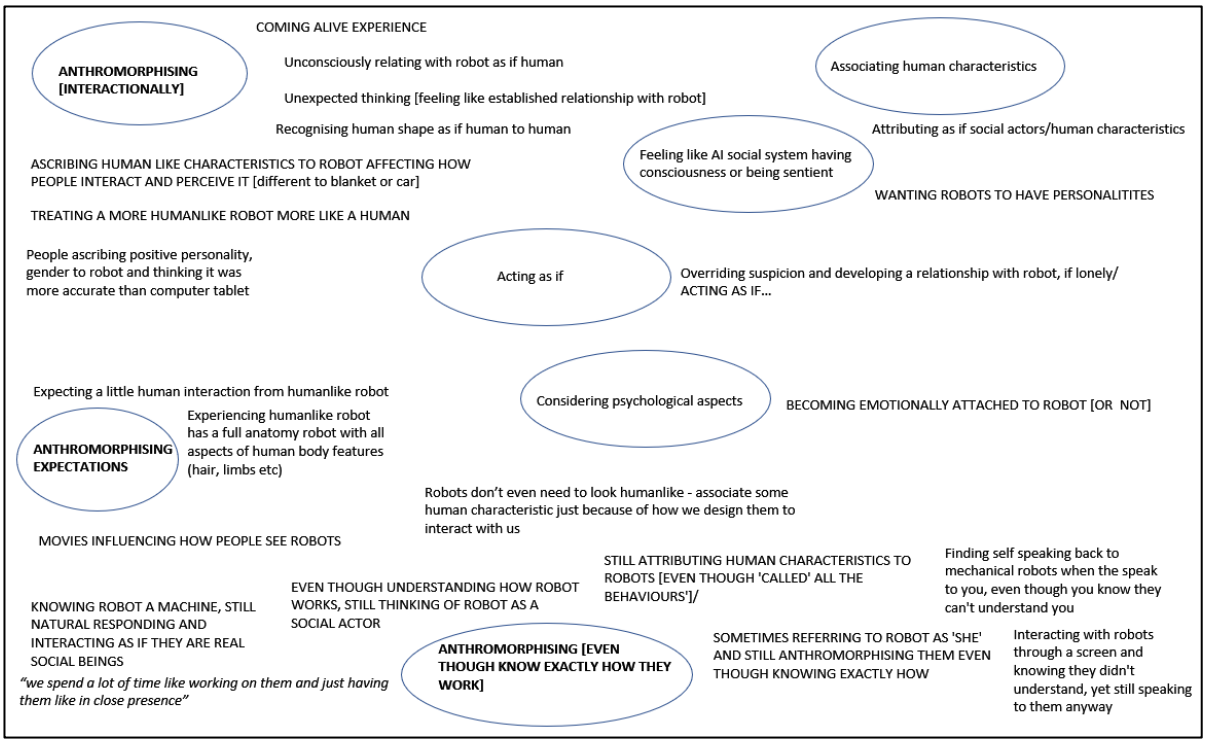
> *Interviewer: With regards to [AI social system], there was a reference to the ability to build relationships. Please can you elaborate and explain?*
>
> *Participant J: people get absorbed into the conversation. Even cynical people forget where they are and what they are doing, somewhere between suspending disbelief that a system could have such a sophisticated conversation and enjoying the feeling of being in a relationship with an 'other' in the anthropological sense of the term.*

As notions and insights developed throughout the study, I compared incidents with earlier abstractions of incidents. As a result, I could identify properties of emerging concepts or potential categories and use the properties for coding incidents that were similar. Memo-writing also formed part of the development of these concepts (Charmaz, 2014); see Figure 5.3 for an example. The grounded theory method contains 'correctives', such as line-by-line coding; thus, the possibility of a researcher superimposing their presuppositions is decreased (Charmaz, 2014). The benefits of conducting initial coding by using line-by-line or incident-with-incident is that it contributes to meeting two criteria for completing grounded theory analysis: fit and relevance. As I have constructed codes and advanced them to categories that illuminate participants' experiences, my research can be said to 'fit' the empirical world. Furthermore, this study has 'relevance' as it is an analytical perspective that provides an interpretation of what was happening and human beings' interactional experiences, elucidating implicit processes and structures in, and as a result of, these interactions (Charmaz, 2014).

Staying close to the data, especially in initial coding, kept me from imposing my presuppositions onto the participants and the data generation; thus, allowing me to consider the material in a new light. It also allowed my prior knowledge and experience into the analysis, assisting in interpreting participants' actions and experiences in the context of larger social processes, as part of developing the grounded theory (Charmaz, 2014). Initial coding enabled me to reflect in depth on the 'contents and nuances' of my data and identify how my prior knowledge shaped my thinking and influenced my coding (Saldana, 2013). I used memo-writing as a means of identifying how my tacit knowledge and presuppositions influenced how I analysed and abstracted data in constructing a theory. Initial coding continued until tentative categories began to form.

Figure 5.3

Example of Memoing: Subcategory 'Anthropomorphising' and It's Properties



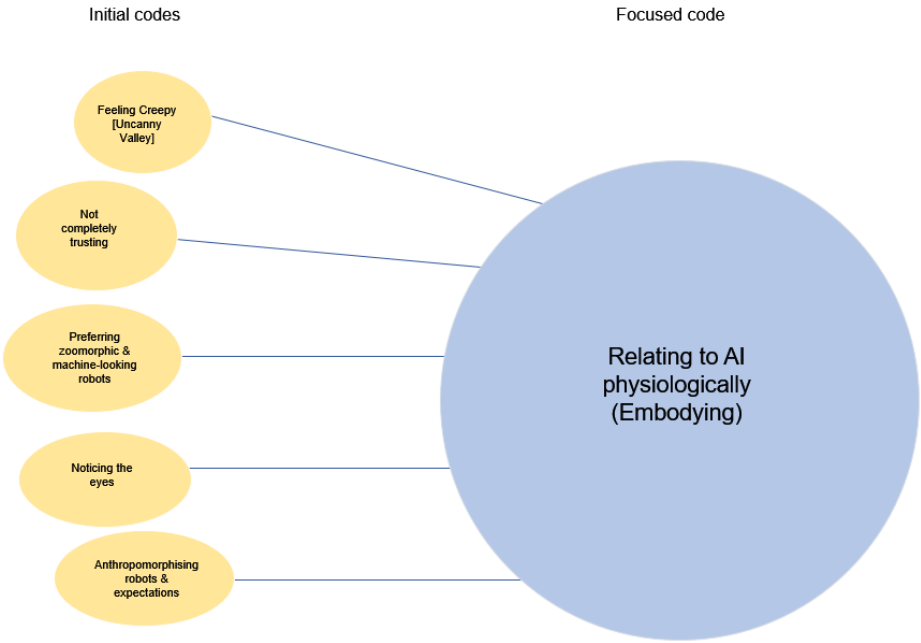
Focused Coding.

In initial coding, I determined significant initial codes that had the most analytical meaning/interpretation and used these focused codes as a means of examining large amounts of data (Charmaz, 2006; Thornberg & Charmaz, 2014). The next phase of coding, focused coding, builds and develops initial coding into more abstract concepts and theoretical integration begins. Through this process of analysis, categories are reviewed and refined until theoretical saturation to identify a core category and theory generation. Focused codes are more directed, capturing and synthesising initial codes into conceptual codes (Charmaz, 2006).

Thornberg and Charmaz (2012) argued that for focused coding, the constructivist grounded theory approach offers more flexibility "by being open for more than one significant or frequent initial code in order to conduct this further work. Such openness also means that the researcher continues to determine the adequacy of those codes during the focused coding" (p. 48). Consequently, I continued to be sensitive and open to adapting my focused codes and to being surprised by the data (Thornberg & Charmaz, 2014). I constructed the focused code '**Relating to AI Physiologically**' (see memo, Figure 5.4) through the constant comparison of initial codes like 'Feeling Creepy', 'Not completely Trusting', 'Preferring Zoomorphic & Machine-looking Robots', 'Noticing the Eyes', and 'Anthropomorphising Robots & Expectations'.

Figure 5.4

Example of a Key Focused Code



During focused coding (see Appendix K) I investigated and decided which codes best captured what I was observing happening in the data. Raising codes to tentative conceptual categories resulted in giving the categories (**Unconscious Feeling**, **Conceptualising**, and **Fantasy to Reality**) conceptual definitions and assessing relationships between them (Charmaz, 2006). Through the generating and refining of categories, I made constant comparisons between codes, emerging categories, and different incidents; as well as compared data from the similar phenomenon in different situations. Forms of comparing also included comparing specific data with the criteria for the category (Thornberg & Charmaz, 2014).

One purpose of focused coding is to ascertain the adequacy and conceptual robustness of initial codes, which includes comparing initial codes with data and identifying the codes that have more analytic power. Furthermore, by comparing codes with codes I could advance my analytic direction and illuminate the theoretical centrality of particular notions or concepts. Charmaz (2014) suggested reflecting on the following list to assist in determining which codes would best make focused codes:

> What do you find when you compare your initial codes with data? In which ways might your initial codes reveal patterns? Which of these codes best account for the data? Have you raised these codes to focused codes? What do your comparisons between codes indicate? Do your focused codes reveal gaps in the data? (pp. 140-141)

I allowed myself to advance the analytical level of one or more codes, as data indicated. I checked how and to what degree the codes fitted other data and whether it led me to asking new questions and gathering further data; thus, determining the usefulness of the code. Initial coding and focused coding are not necessarily a linear process; as such, earlier on I identified initial codes such as 'noticing the eyes'. However, while there were numerous related codes originating from similar experiences, I discerned it did not reflect all that was happening. I asked myself 'what else was happening?' and subsequently discerned the focused code **Relating to AI via embodying physiologically (Embodying**) (see Figure 5.4). This focused code led me to return to earlier participant transcripts, examine topics, and compare codes that were initially too implicit to determine. Upon refining codes, I realised that **Relating to AI via embodying physiologically (Embodying**) was a key focused code that explicated earlier initial codes, thus having greater theoretical reach and centrality. I conducted each subsequent interview and used theoretical sampling and theoretical sensitivity to make explicit what I had sensed but had not been able to capture. Thus, by interacting with components of data, I came to understand what participants were saying and saw the data in a new light, as a considerable theoretical insight. The data contained many statements that fit my focused code relating to **Relating to AI via embodying physiologically (Embodying)** forming an explicit analysis of this central part of the human interactional process.

By abstracting further, I constructed one of the core theoretical categories, '**Unconscious Feeling**', which was a breakthrough juncture. I had an 'aha' moment, recognising that this explained many initial codes and described part of the process of humans' interactional experiences with AI robots or avatars. Consistent with Charmaz' (2014) focused coding, my focused codes followed on naturally from my initial coding, as the successive levels of conceptual analysis were elevated.

Hence, through concurrent comparisons and analysis of significant initial codes and memos, I was able to identify patterns and relationships within and between categories and subcategories, moving to medium-level concepts as I was building my grounded theory. A central aspect of focused coding is the connecting together or integrating of categories.

Theoretical Coding.

Birks and Mills (2015) explained that "theoretical codes are advanced abstractions that provide an outline for increasing the explanatory power of your storyline and its potential as theory" (p. 119). This final phase of coding looked to the descriptive accounts of the processes and interactions between the categories and the core category (Corbin & Strauss, 2008), with the aim of moving the "analytic story in a theoretical direction" (Charmaz, 2014, p. 150). I employed theoretical coding by analysing how the constructed categories and codes from the data correlated and formed hypotheses that could then be integrated into a substantive theory. Theoretical codes served as analytical tools for me to arrange and conceptualise the developed codes and categories in order to build a comprehensive grounded theory (Glaser, 1978). Thornberg and Charmaz (2014) noted that initial and focused coding are derived from codes and categories that are data-driven and empirical as a result of constant comparative analysis of data, data and codes, and codes and codes. Theoretical codes, however, are derived from insights and views that a researcher brings into the research process as 'analytic tools and lenses' from varying external, pre-existing theories in the form of concepts, terms, or abstract models that "specify possible relationships between categories [that a researcher has]… developed in [the] focused coding … [and] may help … tell an analytic story that has coherence"' (Charmaz, 2006, p. 63).

Thornberg and Charmaz (2014) recommended researchers examine all types of extant theories from various research disciplines so as to discover the theoretical codes within for themselves. Consequently, theoretical codes were imported consciously and with reflexivity. An accumulation of numerous theoretical codes tends to highlight the relationships between theoretical categories and is utilised when connecting and arranging categories and incorporating them into a grounded theory. Consequently, I used constant comparisons between data, theoretical codes, analytically generated codes, and categories, as well as memos. I worked in line with Glaser's (1978) argument that 'theoretical codes must earn their way into analysis'; thus, having relevance and fit the data, as well as developed and refined categories. For example, I took advantage of many theoretical codes to develop my categories and examine their relationships to each other in order to integrate them into my grounded theory, such as the theoretical code '**Embodied**'.

Taking into account Glaser's (2005) suggestion that theoretical codes can be used from other disciplines and that more than one theoretical code can be applicable in the construction of a grounded theory, I applied theoretical constructs from my own discipline of psychotherapy, as well as other disciplines, such as neuroscience (Damasio, 2003; Iacoboni, 2009) and cognitive science and philosophy (Lakoff & Johnson, 1999); thus, extending, supporting, and validating existing theories while explicating and strengthening the merit of my own contribution. Further, the storyline and theoretical coding techniques functioned as a conduit between analysis and theory (Birks & Mills, 2015). According to Thornberg and Charmaz (2014), abductive reasoning is used when conducting theoretical coding, as researchers examine their accumulative theoretical codes and compare them with their own research data, developed codes, and categories. As such theoretical codes move the analysis in a theoretical direction, underpin substantive codes and demonstrate correlations between them. Furthermore, by explicating experience, codes establish a link between defined data and the researcher's developing analysis (Charmaz, 2014).

Charmaz (2014) highlighted that theoretical sensitivity and coding shape each other, with theoretical sensitivity demonstrating understanding and correlations of the studied phenomena as well as explicating the phenomena in abstract terms. By employing theoretical sensitivity, I was encouraged to go to the core of human interactional experiences, as the studied phenomena, and identify how it was comprised and what it entailed (Charmaz, 2014).

Software

I used computer software to help manage the data and facilitate the coding process (Bringer et al., 2004; Hutchison et al., 2010). For the first three interviews, I inserted each transcript into its own Microsoft Excel worksheet and added two narrower columns to the right; one for initial line-by-line coding, and the other column having the gerund for each initial code (see Appendix K example of interview data and initial coding). In the initial coding stages, I used Microsoft Excel and then NVivo software to organise, access, and manage the data. The software allowed me to organise numerous initial codes and, specifically in NVivo, I could link the codes directly to the raw data in the interview transcript. During the focused coding stages, I used Microsoft Excel as I found it useful when creating and collapsing categories and subcategories. I also used Microsoft PowerPoint to visualise the focused codes which helped me to helicopter above the data as I engaged in abductive reasoning. As I moved from focused coding to theoretical coding, I continued to utilise Microsoft Excel and PowerPoint as my main tools for comparisons and analysis.

Memo Writing

Memo writing is a crucial part of grounded theory research analysis. Memoing my contemplations informed my decisions about sampling, coding, categories, and theoretical relationships within the analysis, providing additional explanations about the grounded theory findings (Birks & Mills, 2015; Charmaz, 2014). My memos contained planned activities as well as changes in direction and rationale for various choices and responses to data that arose during the study. Additionally, I explored the meanings and consequences of occurrences throughout the research process. As I worked with the data, I utilised my memos to clarify, explore, and examine the interpretations I drew on (Birks et al., 2008); thus, raising theoretical sensitivity. Through memoing, I interrogated the data with the purpose of creating abstract concepts required for constructing a theory and detecting potential gaps or deficiencies in my analytical thinking earlier in the research process (Corbin & Strauss, 2008).

Memo-writing fosters the analysis of the research data and codes from early on in the research process, and keeps the researcher involved in the analytic work through successive memos (Charmaz, 2014). Birks and Mills (2015) referred to memoing as "an uninhibited activity in which you are free to explore your ideas, instincts and intuition in relation to your research" (p. 40). My memo-writing was spontaneous and informal, "preserving [my] natural voice" (Charmaz, 2014, p. 181). I had started hand-written memos in notebooks; however, I transferred these memos into soft copy on Microsoft PowerPoint, dating each memo and creating a new slide for any amendments or changes creating a 'memo bank' (Clarke, 2005). This process proved to be useful as I could then search the entire memo document with ease. As my comfort in memo-writing developed it increased my reflexivity, depth, and richness of content by fostering questions on my presuppositions and potential shaping of my analytical interpretations (Birks & Mills, 2015; Charmaz, 2014; Strauss & Corbin, 1998). My memos ranged from personal reflections and assumptions about my research, wonderings, insights, and potential issues (Birks & Mills, 2015). I also included diagramming, as well as a photo (see Appendix P). In my memos, I considered questions such as:

> What process is at issue here? How would [I] define it? To what extent is it explicit or does it remain implicit? Under which conditions does this process develop? How do(es) the research participant(s) think, feel, and act while involved in this process? What slows, impedes, or accelerates the process? What are the consequences of the process? (Charmaz, 2014, pp. 169-170)

Memoing throughout the research process assisted me in examining and exploring my codes and categories, including any correlations between them. Consequently, I was able to advance the level of abstraction of my insights and ideas.

I used memos to develop focused codes into tentative conceptual categories (Charmaz, 2014) and through subsequent sorting, modifying, and exploring of my memos I often wrote further memos that facilitated theory integration and theory construction (Birks & Mills, 2015; Charmaz, 2015). Indeed, “memo-writing is the methodological link, the distillation process, through which the researcher transforms data into theory” (Lempert, 2007, p. 345). Examples of memos are in Appendices M, N, and O. I was mindful to memo my observations and insights, even if they differed from the statements in the data. I considered if these notions relied on obscured meanings and actions that had not yet completely emerged by focusing on making analytic sense of the substantive ideas while addressing taken-for-granted assumptions. Thus, data analysis, in part, depended on my prior knowledge and assumptions, while being open to viewing these as one of many perspectives.

For example, as highlighted in my presupposition interview, I used psychotherapeutic concepts, such as attachment theory, through which I usually explain aspects of human relating and interactions. I aimed to avoid prejudging what was happening and assuming participants’ attachment styles. Rather, I focused on understanding participants’ experiences and actions, thus aiming to understand their reality and reasonings of their experiences so as to gain new insights. Subsequently, if I used terms from my professional knowledge as codes, I did so more consciously and ensured that they ‘fit the data’ (Charmaz, 2014).

Categorising

Through the process of constant comparative analysis, I developed multi-dimensional categories with several subcategories that, combined, explicate the constructed grounded theory (Birks & Mills, 2015). Furthermore, these categories and subcategories include properties which are determined by their dimensions and connect to the conditions under which they operate (Birks & Mills, 2015). Techniques, such as memo-writing, which raise theoretical sensitivity, can advance the development of properties and dimensions of categories and subcategories (Birks & Mills, 2015). As I refined categories by defining properties and their dimensions, while taking into account the conditions under which they operated, I was able to recognise gaps in the data and use theoretical sampling until I reached theoretical sufficiency (Dey, 1999). Further to Morse’s (1995) statement that in grounded theory “researchers cease data collection when they have enough data to build a comprehensive and convincing theory” (p. 148), I chose not to conduct member checking. Indeed, both Birks and Mills (2015) and Charmaz (2006) concurred that it was ‘inherently unreliable’ as participants’ responses were within the context at the time of the interview. Additionally, in grounded theory method member checking is “redundant as a source of verification for conceptual analysis” (Birks & Mills,

2015, p. 97). Consistent with Charmaz (2014), rather than identifying a core category (Glaser, 1978; Strauss, 1987; Strauss & Corbin, 1990), I took a wider approach by explicating how the categories and their subcategories integrate and correlate to form a substantive grounded theory.

Diagramming

As my mode of learning is more visual, I attempted to diagram early on in the research process to gain a visual picture of what I was coding. However, I discovered I was somewhat overeager, getting ahead of myself, as I had not developed the conceptual level of abstraction to a point that a diagram would make sense. I did though start to consider the flow and movement of the categories and subcategories in terms of the processes and actions of the developing grounded theory. Consequently, diagramming, while simultaneously generating data, assisted in mapping and linking different codes, and further conceptual analysis (see Appendix P) as a form of abductive reasoning. As the categories and subcategories became more substantive, I wanted to express the movement and dynamics of the human being and how their experience shaped their interactional process. This is when I identified the meaning and the movement in the human DNA helix as an expression of the constructed grounded theory, by distinguishing concepts and illustrating relationships. Furthermore, the recursive process of visual thinking, through diagramming, helped to identify gaps in my developing theory which led to more theoretical sampling; and to notice points of connection and overlap, fostering creative advances in interpretation and in-depth understanding of ideas and data (Charmaz, 2014; Kennedy & Thornberg, 2018; Strübing, 2007). For example, I considered the subcategories ‘Getting in Flow State’ (Properties: Co-creating experiences; Becoming more real; No jarring); ‘Being Lulled’ (Properties: Being taken in; Keeping in the moment; Ignoring details & focusing on how feeling); and ‘Forgetting’ (Properties: Being in the moment; Not realising; Snapping out of the magic) and whether they overlapped or were distinct. However, the properties highlighted distinct experiences, allowing me to make more in-depth interpretations which led to illuminating a key aspect in the process.

Diagramming allowed me to examine, test, modify, and integrate concepts while building theory (Birks & Mills, 2015; Charmaz, 2014; Corbin & Strauss, 2015). While my grounded theory is line with Charmaz’ (2014) constructivist grounded theory, I did consider and analyse contextual conditions, to the degree that they addressed the second part of the research question. Reflecting on my philosophical lens of pragmatism and symbolic interactionism, I resonated with the following quote that: “while actors act towards their environment they are both socially and materially a part of their surroundings… subject, cognition, knowledge and theory all need to be conceptualised as a dynamic process

resulting from activity relating to actors with environment" (Strübing, 2007, p. 595). As a result, in diagramming I included the third category 'Fantasy becoming Reality' as part of the human being's interactional experience shaping and being shaped by contextual conditions.

Storyline

The storyline technique, defined as "the conceptualisation of the story… the core category" (Strauss & Corbin, 1990, p. 116) can aid grounded theory analysis. The storyline explains the grounded theory and the theoretical constructs that make up the theory and advances the conceptual level of analysis. A benefit to using a storyline is that any gaps are identifiable in the developing theory. Charmaz (2006) noted that sometimes cases do not appear to fit the overarching theoretical concept and these negative cases come straight from the data. These variations promote refinement and add depth to the emerging theory. Furthermore, Strauss and Corbin (1998) purported that including variation into the end product of the grounded theory advances its scope and illuminating potential. Such was the case in my study, where I conducted a storyline to assist in identifying any gaps and variations, specifically in relation to whether and when people disengaged or not with AI social systems, and how data that seemed important did not appear to fit the overall theoretical concept. For example, Participant I reported not moving past the subcategories of '**Buying into the Hype**' and '**Having Expectations**', thus disengaging and not having an '**Embodied Relating**' experience with the AI social system. Charmaz (2014) contended that "being selective about which data you seek and where you seek them aids you to discern variation in your category or studied process" (p. 207). In order to determine why my theoretical categories varied, I questioned 'how can I account for this phenomenon?' (Charmaz, 2014). Memoing and further theoretical sampling led to me identifying the third theoretical category '**Fantasy Becoming Reality**'. Furthermore, by going back through the data, and employing theoretical sampling to refine and modify analysis and advance the depth of the first two categories, the process of disengagement or continued engagement within the interactional experience to explain this variation was indicated.

Advanced Analysis

The interplay between initial and focused coding, simultaneous data collection, or generation and comparative analysis with theoretical sampling is an iterative process. The early phases of coding produced a large number of codes and potential categories. At one point I became stuck, producing a memo asking, 'And so what?' Consequently, I returned to my data, codes, and categories, and used theoretical sensitivity and further theoretical sampling to move from medium to high-level concepts.

The subcategories of the three theoretical categories were developed largely from participants' accounts in order to stay close to the data. Birks and Mills (2015) advocated that theoretical integration requires "the application of advanced analytical strategies in order to raise your analysis to the highest conceptual level possible" (p. 109). By raising my theoretical sensitivity and scrutinising categories before asserting that I had reached theoretical sufficiency, I could then present an integrated theory grounded in the data. Charmaz (2014) took an interpretivist perspective of theory, viewing it as more abstract and indeterminate. Consistent with the essential grounded theory methods, I used memos for documenting my analytical ideas and as a tool throughout my research process. For instance, during theory integration I sorted my memos to assist in the final construction of my grounded theory (Birks & Mills, 2015) by determining correlations and synthesising concepts that were not evident earlier. As suggested by Birks and Mills, both the storyline technique and theoretical coding are notable strategies that can advance theoretical integration. The constructing of a grounded theory and integrating the theory, while challenging, resulted in an original theory that is intended to contribute to knowledge within my own profession and beyond.

Theoretical Sufficiency

Charmaz (2014) stated that the core aim of theoretical sampling is to explicate and advance the categories forming the grounded theory, and to foster the properties of the categories until no new properties arise. Winer (2007) defined theoretical saturation as a "judgment that there is no need to collect further data" (p. 306). Due to the emergent nature of grounded theory, the notion of 'theoretical sufficiency' can be applied as a guideline for establishing sample size and when the generation of data is sufficient for theory construction (Charmaz, 2014; Stern, 2017). While many grounded theorists (Birks & Mills, 2015; Corbin & Strauss, 2008, 2015; Glaser & Strauss, 1967; Strauss & Corbin, 1990) include the concept of 'theoretical saturation', others (Dey, 1999; Low, 2019) have suggested that the concept is 'imprecise' and problematic. Charmaz (2014), herself, queried the validity of assertions of saturation. Dey (1999) argued that instead of determining categories saturated by data, categories can be 'suggested by data'; and posited that in terms of theoretical saturation the focus should be of quality not quantity. As such, rather than achieving saturation, Dey maintained that the concept "theoretical sufficient" (p. 257) was a more suitable fit for grounded theory researchers. Furthermore, Dey emphasised that the saturation of data should not be at the cost of the advancement of categories. I implemented the concept of theoretical sufficiency when my data and analysis did not achieve any further theoretical explicated detail of the categories, subcategories, and the constructed theory (Charmaz, 2014). Therefore, considering Low's (2019) suggestion of reaching 'pragmatic saturation', I defined, verified, and

explicated correlations between categories and properties within these categories, establishing patterns that were insightful and showing analytic rigour and identify abstract theoretical links (Charmaz, 2014). Analysis of the grounded theory constructed during this study is presented in the Discussion, Chapter 10.

Rigour

Through the process of analysis, the aim was to build a theory that explained processes and actions, with the hope that the theory may be useful in other contexts, and "still retain its grab and fit" (Bryant, 2009, p. 25). Grounded theory is considered one of the most disciplined and rigorous of the qualitative approaches, where procedural precision upholds the quality and rigour of the study through the vigilant management of data resources, as well as an audit trail that shows the analysis and process (including memos) of the research that resulted in the final grounded theory (Tie et al., 2019). In addition, the quality of study emphasised the validation of methodological congruence of my philosophical position with the research question and chosen methodological approach (Tie et al., 2019). As such, by rigorously applying the essential grounded theory methods and attentively employing memo-writing, I strove to ensure quality in both the processes and outcome of the grounded theory research (Birks & Mills, 2015).

Evaluation of Grounded Theory

Charmaz (2014) considered the criteria for evaluating research varies depending on who develops them and for what purpose(s). In terms of evaluating grounded theories, even though they can vary, Charmaz provided the following criteria: credibility, originality, resonance, and usefulness. **Credibility** is concerned with the criteria of having developed an in-depth familiarity with the topic; and adequate, applicable data in order to progress with insightful questions regarding the data, while making constant comparisons throughout the study and linking data collected with the detailed analysis. Thus, credibility is about providing sufficient evidence to support the researcher's claims and for independent assessment (Charmaz & Thornberg, 2020). **Originality** pertains to the importance of providing an analysis that has theoretical significance, extending current grounded theory concepts and offering new insights (Birks & Mills, 2015). **Resonance** refers to the reflected depth of the analysis and constructed theory that represents, fully, the participants' experiences, revealing deeper meanings and insights, and providing a grounded theory that makes sense to participants and others who share their situations (Charmaz, 2014). **Usefulness** relates to the applicability of the analysis and interpretations to everyday situations, considering the generic processes as a result of the analysis and any implied consequences.

Furthermore, consideration is given to whether the analysis contributes to further research and knowledge in other areas (Charmaz, 2014). Additionally, Lincoln and Guba's (1985) evaluative notion of transferability, where practitioners can apply the knowledge from the study and provide insights into the contexts and meanings for the people for whom they work, means the researcher is required to provide a detailed and comprehensive analysis and theory in order for it to be transferable to practitioners. Charmaz argued that:

> The constructivist approach fosters renewal and revitalization of grounded theory by integrating methodological developments with the original statement of the method. This approach challenges the assumption of creating general abstract theories and leads us to situated knowledge…, while simultaneously moving grounded theory further into interpretive social science. (p. 241)

A robust synthesis of Charmaz's (2014) evaluating criteria of originality and credibility enhances resonance, usefulness, and the ensuing value of the contribution. Charmaz (2006) suggested that "a grounded theory that conceptualizes and conveys what is meaningful about a substantive area can make a valuable contribution" (p. 182).

Discussing the Theory

Following constructivist grounded theory (Charmaz, 2014), through the interview process, participants co-constructed data. As such, their views and experiences are represented throughout the discussion of the findings. When describing the theory in terms of humans' processes and experiences in the subsequent Findings chapters, I stay close to the data by preserving the participants' voices at the forefront.

Throughout the thesis, when I have quoted a participant, it is noted in italics, along with their pseudonym (see Table 1). Quotes are verbatim, except for minor editing for articulation such as adding [square brackets] to participant quotes. I also removed words or small phrases such as 'like', 'um', or 'you know' which do not lessen the intrinsic meaning in the quote. Occasionally, I have used one participant's description or example to depict the intrinsic principle of numerous viewpoints to provide a coherent account of a component of the theory.

Summary

This chapter has outlined the grounded theory research methods used regarding the stages of analysis and theory construction. Data analysis was shaped by the theoretical content and direction of the study which influenced the amount and type of data gathered and collected, adopting the iterative process of constant comparative analytic strategies in coding, memo writing, sampling, and integrating the analysis (Charmaz, 2008).

I have described how the use of grounded theory methods led to the constructed theory as laid out in the subsequent Findings chapters. This study sought to investigate human beings' experiences of interacting with AI robot and/or avatars, explicate a process relating to the research question, and generate a theory grounded in the data.

In the following chapters the research findings are presented. In Chapter 6, the theory of **Embodied Relating** is outlined. The subsequent three Chapters 7 to 9, and Discussion, Chapter 10, consider each of the theoretical categories and their subcategories: **Unconscious Feeling, Conceptualising**, and **Fantasy Becoming Reality**, which encompass the constructed theory of **Embodied Relating**.

The mind is inherently embodied.
Thought is mostly unconscious.
Abstract concepts are largely metaphorical.
(Lakoff & Johnson, 1999, p.3)

Chapter 6: Theory of Embodied Relating

Introduction

Using the grounded theory research process, the study aimed to answer the following research questions: 'How do human beings experience AI social robots and/or digital avatars in an interactional context?' 'What is happening and how will it inform the future?' This chapter explains the findings of the constructed theory, as a storyline, that has been grounded in the data. The resultant theory, **Embodied Relating: Thinking with feelings**, answers these questions and is presented by explaining the three theoretical categories: **Unconscious Feeling (Visceral responding)**, **Conceptualising (Constructing of an 'other')**, and **Fantasy Becoming Reality**, and how human beings experienced **Embodied Relating**.

Introducing the Theory of Embodied Relating

Table 6.1 below provides an overview of the theory of **Embodied Relating**, followed by an outline of each of the three theoretical categories and their subcategories. Presenting the overall theory and its three main categories as a storyline is a means of explaining the theory to the reader, and forms part of the framework in grounded theory for advanced analysis wherein categories are linked and gaps during theory integration recognised. This chapter outlines the overall theory constructed from the research, with each of the three categories subsequently represented in their own chapters, explicating an in-depth account of each of the three theoretical categories and their corresponding subcategories and properties. The theory of **Embodied Relating** was constructed using grounded theory processes to develop a core category with three linking categories and their subcategories (Charmaz, 2014). It is a dynamic theory that explains how human beings experience interacting with AI robots and avatars, which feeds into human beings' developmental and innate processes and actions of relating to others socially, both genetically and socially influencing these interactions. Key findings are that **Embodied Relating** is an unconscious, visceral responding, **Unconscious Feeling**, as well as a conceptual understanding and meaning-making experience, **Conceptualising**, with the interactional experience varying depending on how the subcategories are impacted in the interaction with an AI robot and/or avatar. Subsequently, the third category **Fantasy Becoming Reality** explicates how each of these experiences shapes current and future experiences and how society influences perceptions of human beings.

A fundamental assumption of constructivist grounded theory is to give voice to participants. Therefore, Charmaz (2006) encouraged grounded theorists to integrate the multitude of voices, viewpoints and concepts of participants in representing their lived experiences. As a constructivist grounded theory methodology, this research focused on the interpretive understandings of participants' meanings and the findings were grounded in the data, therefore, citations from other authors were not included in the Findings chapters. Participants' comments support the storyline and evidence the construction of the categories and the theory. Where feasible, I have preserved the verbatim nature of participants' words to retain the integrity of what they reported and to show analytical rigour. Insertions to assist clarity are denoted by […]. Un-bracketed ellipses, … denote omitted text.

Table 6.1

Outline of the Findings Chapters

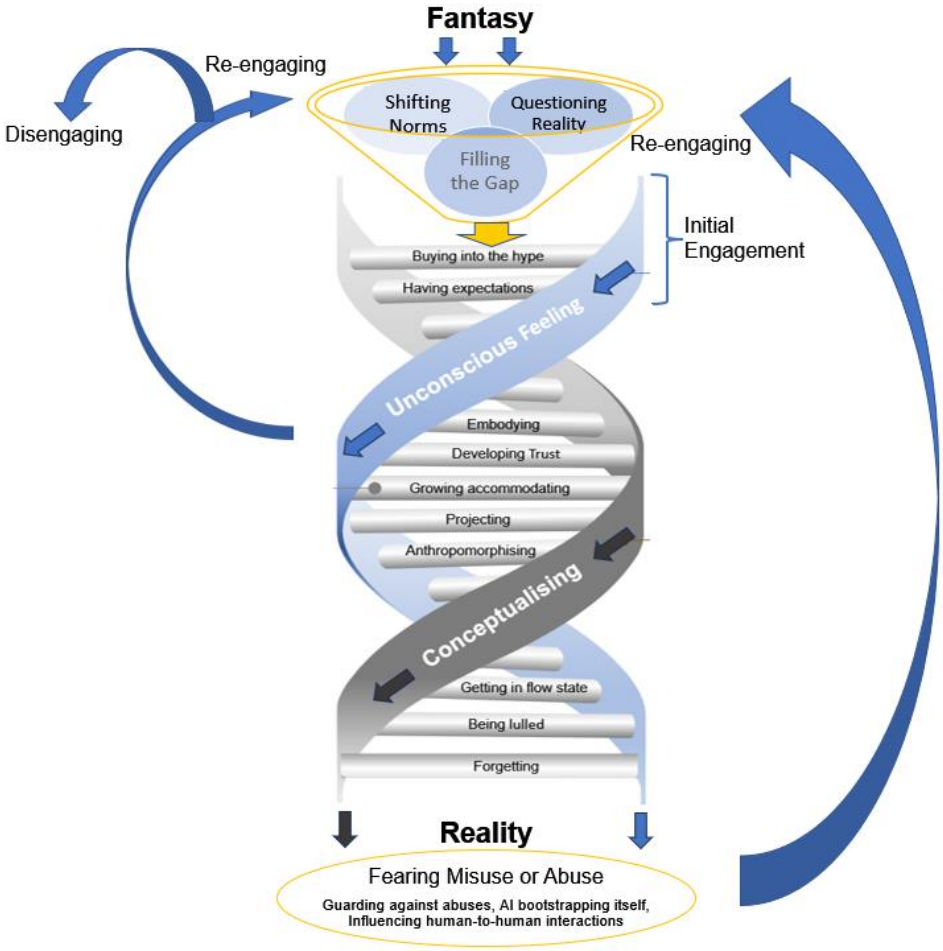
Chapter	Content
Chapter 6	**Core category: Embodied Relating: Thinking with feelings** Overview of the grounded theory, with three theoretical categories and their subcategories, properties, and dimensions
Chapter 7	**Theoretical Category: Unconscious Feeling (Visceral responding)** Subcategories: **Buying into the Hype** (Properties: Taking the bait; Being pushed; Feeling honoured) **Having Expectations** (Properties: Being influenced by science-fiction; Being influence by mass media; Eliciting expectations depending on appearance) **Embodying** (Properties: Engaging viscerally; Noticing the eyes; Experiencing facial expressions and non-verbal movements; Feeling physical versus digital embodiment) **Developing Trust** (Properties: Lacking believability and trust; Not completely trusting; Building trust using humour; Building trust through repeated positive interactions) **Growing Accommodating** (Properties: Feelings of tenderness; Persisting in wanting to make a connection; Getting used to; Forgiving; Suspending disbelief)
Chapter 8	**Theoretical Category: Conceptualising (Constructing an 'other')** Subcategories: **Projecting** (Properties: Making real sense; Perceiving personally motivated) **Anthropomorphising** (Properties: Associating human characteristics; Evolving autonomy; Considering psychological aspects; acting as if; thinking things as having consciousness or being sentient) **Getting in Flow State** (Properties: Co-creating experiences; Becoming more real; No Jarring) **Being Lulled** (Properties: Being taken in; Keeping in the moment; Ignoring details & focusing on how feeling) **Forgetting** (Properties: Being in the moment; Not realising; Snapping out of the magic)
Chapter 9	**Theoretical Category: Fantasy Becoming Reality** Subcategories: **Reality** (Properties: Straddling a strange line; feeling [as if AI social system] alive and conscious) **Shifting Norms** (Properties: Generational shifting; Speeding-up from Covid-19 pandemic; Using social media and gaming; Culture playing a role) **Filling the Gap** (Properties: Meeting a need) **Fearing Misuse or Abuse** (Properties: Guarding against abuses, AI bootstrapping itself; Concerning uncertainty & affects on human-to-human interactions)

Explanation of the Diagrammatic Representation

To provide further understanding of the **Embodied Relating** process, Figure 6.1 depicts a double helix diagram, which provides a visual explanation of the interactional experience with AI robots and avatars as described by human beings. The three theoretical categories making up this theory are: **Unconscious Feeling, Conceptualising**, and **Fantasy Becoming Reality**. The two theoretical categories that address the first research question (**Unconscious Feeling** and **Conceptualising)** do not occur in a linear process; instead, the two categories continually occur in conjunction with their respective subcategories: **Buying into the Hype**, **Having Expectations**, **Embodying**, **Developing Trust**, **Growing Accommodating**, **Projecting**, **Anthropomorphising**, **Getting in Flow State**, **Being Lulled**, and **Forgetting**.

Figure 6.1

Theory of Embodied Relating



Although the subcategories are concurrently shaped and shaping the theoretical categories dynamically, they move progressively through a set process that is reflected in Figure 6.1, where the human beings' interaction from beginning (engaging) to the end

of the interaction (disengaging) are demonstrated by arrows on the helix. The blue arrows show whether they sought to re-engage or whether they disengaged, depending on whether the human-user felt the interactional experience was positive and felt good or not. While these categories are inextricably linked and these subcategories connect to both theoretical categories influencing them and being influenced by them throughout the interaction, they have been discussed separately for explanatory purposes. The categories of Unconscious Feeling and Conceptualising, and their subcategories, continually impacts and is impacted by the third category, **Fantasy Becoming Reality**, as it moves between the individual to society and back again. It is a continually moving and changing process, with each category dynamically influencing and shaping each other. The subcategories **Reality**, **Shifting Norms**, and **Filling the Gap** all influence a human being's interactional experience with an AI social robot and/or digital avatar, with the outcome shaping their reality. This reality feeds back into the fantasy and again contributes to the shaping of the human being's experience interacting with further AI robots and/or avatars. An outcome of this reality is the subcategory **Fearing Misuse and Abuse** (which includes 'guarding against abuse, AI bootstrapping itself, and influencing human-to-human interactions'). The helix diagram demonstrates the process of the two categories **Unconscious Feeling** and **Conceptualising** (and their subcategories) within the individual human-user's interactional experience with an AI social system. As well as demonstrating how, in conjunction with the third category which includes dynamics external from the individual human-user's experience, together they shape each other, in a constant evolutionary dynamic.

The use of the helix in the diagram is also representative of both the biological and conceptual processes of human beings that is continually developing and evolving, and is being influenced by and is influencing society's evolvement. This theory is dynamic and has flexibility in its application and understanding of individual interactional experiences (micro level), as well as at a societal experience and process (macro level).

Unconscious Feeling: Thinking with Feelings

This theoretical category explains the visceral responses human beings have when socially interacting, pointing to human beings' emotions being intrinsic to human communication, social interactions, and cognition, and unconsciously attending to how they feel in the interaction. There are five subcategories: **Buying into the Hype**, **Having Expectations**, **Embodying**, **Developing Trust**, **Growing Accommodating**, which form the first processes within the person's interactional experience with an AI robot and/or avatar. It is at this point that if the person does not grow accommodating, they disengage from the interaction. This is a key point in the interactional process of continuing to engage or disengage.

Buying into the Hype

This is where the process of choosing to engage with the AI robots and/or avatars begins, usually as a result of these technologies being marketed as relational, helpful to human beings, and being more advanced than they actually are, in order to entice people to buy and use these products. The "*marketing hype*" (Participant J) of these AI social systems has been referred to as people "*being pushed*" (Participant K) towards this technology. When participants were presented with the opportunity to interact with an AI robot and/or avatar, most seemed to be curious and excited, leading to many participants feeling "*honoured*" (Participant B) to interact with these technologies. Even when people were suspicious of these AI social systems and not keen or ready to buy into the hype and engage with an AI robot and/or avatar, some companies buy into the hype and purchase AI social systems with the expectation that they will meet a need of theirs (e.g., whether for enhanced customer service or more efficient sales outcomes). Therefore, the dimensions range from complete buying into the hype – seeking the experience out – to those who are more reluctant and suspicious.

Having Expectations

Participants described having higher expectations of the AI robot and/or avatar capabilities than what was the reality of the actual technology which led most participants to become disappointed or disillusioned or left wanting more by the disparity between what they had expected this technology to do and what its actual capabilities were. Science-fiction stories, especially from movies, have influenced people's expectations, with one participant stating, "*a perception or expectation is set by the science fiction* [of] *… how intelligent or how good or how bad these robots can be*" (Participant K). Additionally, participants spoke of gaining considerable expectations from mass media (i.e., social media). Furthermore, when a social robot looked human-like, for example, people expected and thought it looked more helpful and intelligent and could help them. Participants who interacted with an AI robot or avatar (that looked like a human being) expressed their expectations that they could "*do what humans do*" (Participant I).

Embodying

Findings indicate that how an AI social system is embodied changes how human beings interacted and related to them, and their resultant interaction experience. The implication being that this AI is sharing a space with human beings; and, in that sense, 'embodying'. One frequent response was participants noticing the eyes of the AI robot or avatar, as people tend to connect with their eyes when engaging with others. For example, "*I always look people in the eyes when I talk to them*" (Participant A).

When an AI social system is embodied, participants unconsciously take into account the non-verbal movements and facial expressions of human-looking AI robots or avatars. Thus, embodying makes it more accessible for human beings to engage with these technologies. Equally if the AI social system did not quite look like human beings and was not moving quite right, it led to participants getting a "*creepy*" (Participant L) feeling. For example, another participant who interacted with a human-looking digital avatar had an instinctive response wanting to touch the AI avatar, using their senses, yet 'knowing' the AI avatar was on a screen. The participant recognised that they were "*Unconsciously reaching out, wanting to touch the* [AI avatar's] *hair"* (Participant M)

Developing Trust

Human beings' level of trust in AI robots and/or avatars ranged from lacking believability and thus "*not completely trusting*" (Participant B), to how human-like the AI social systems behaved with their use of humour and facial expressions, and "*building trust*" (Participant N).

> *it doesn't mean …I put all my trust in the robot.* (Participant B)
>
> *The trust comes in … if they [AI social systems] …helping them achieve what they wanted to achieve.* (Participant N)

Throughout the interactional process, the level trust can increase or decrease depending on the human being's experience in relation to all the links connected to the two categories (**Unconscious Feeling** and **Conceptualising**), and how they feel about the interaction, within the helix. There are variations of believability and trust when engaging with AI social robots or digital avatars.

Growing Accommodating

In general, human beings referred to AI social robots and/or avatars as "*creatures*" (Participant B), and social robots that looked more like toys or animals as awakening "*feelings of tenderness*" (Participant C). Some participants commented on "*getting used*" (Participant H) the digital avatar or "*growing more accommodating to the* [AI social system's] *quirks*" (Participant N) and "forgiving" (Participant E) the AI social system when it made mistakes. Throughout the interviews, participants persisted with their interactional experience wanting to make a connection with the AI social robot and/or digital avatar, even when it did not respond or behave as expected. Other spoke of "*suspending disbelief*" (Participant I) and "*overriding suspicion in order to develop a relationship*" (Participant C) with the AI social system.

This first theoretical category, **Unconscious Feeling**, highlights findings in the study which point to the unconscious, visceral responses human beings have from before they see the social robot or digital avatar, to the moment they meet, through engagement and the interactional experience that moves through the various subcategories (as discussed above and in more detail in Chapter 8).

Conceptualising: Constructing an 'Other'

While both the theoretical categories **Unconscious Feeling** and **Conceptualising** are central to the entire human beings' interactional experience with an AI robot and/or avatar, the second group of subcategories in the process—**Projecting**, **Anthropomorphising**, **Being Lulled**, **Getting in Flow State**, and **Forgetting**—focus on the conceptualising and constructing of an 'Other'; how it influences the interaction and how the interaction is experienced by the human being.

In this study, as the interaction progresses through the dynamic DNA helix of human-user experience, the human being constructs an 'Other' with the AI robot and/or avatar. For example, "*trying to create a theory of mind with the* [AI social system]" (Participant J); in this instance, through symbolism of shared concepts, language, and body language (non-verbal communication), as well as **unconscious feeling**.

Projecting

Projection is the idea that as a way for human beings to make sense of their experience engaging with an AI social system, part of these perceptions is based on actual interactions with the [AI social system] and part is suggested to be "*purely personal motivations and understandings lying underneath that*" (Participant I). The human being's subjective interactional experience with an AI robot and/or avatar can be affected by their projections onto the AI social system and the connection, as "*a matter of personalisation*" (Participant L), that can change over time. In other words, the AI robot and/or avatar becomes something to the human-user who projects that perception onto it.

Anthropomorphising

Many participants considered the AI robot and/or avatar as a social actor, "*associating human capabilities*" (Participant C) to them. When an AI social system appeared more relational, it contributed further to the association. Throughout the study, participants disclosed, unconsciously or consciously, that they anthropomorphised AI robots and/or avatars. While interacting with these embodied AI social systems, human beings began to assume that these technologies had sentience 'as if' they were a human being. Additionally, if the conversation the human being has with the AI robot and/or avatar feels human-like to the person, then it drives the person to consider psychological aspects of that AI social system, leading to attributing human characteristics to it more easily.

Additionally, the design of socially intelligent robots and digital avatars that have evolving autonomy contribute to how human beings perceive and then anthropomorphise these AI social systems.

Getting in the Flow State

As part of the conversational flow between a human being and an AI social system, there was a sense of a "*co-creating*" (Participant O) or the co-constructing of an experience in "*a constant feedback loop*" (Participant O). As such, the conversational flow continues unless the AI social system does something that is "*jarring*" (Participant C) to disrupt this flow.

> *not mattering if the rest of it doesn't have good graphics…, as long as it is not jarring, then actually start to go with the flow [in conversation] … [so that] the minute the bot says something jarring…then forget about it…the magic disappeared.* (Participant C)

The importance of conversational flow is reflected in Participant C's comment that there was nothing jarring in the conversation interaction, then getting "*to get to the flow state …you need to have nothing…that reminds you that this is artificial*". Part of human beings 'getting in a flow state' when interacting with AI robots and avatars, is that they appear 'as if' becoming real over the course of the interaction due to the AI social systems mimicking human emotions and behaviours, thus creating a more natural, engaging, and flowing interactional experience.

Being Lulled

As the interactional experience progresses through the subcategories, and in conjunction with the theoretical categories **Unconscious Feeling** and **Conceptualising**, people experienced "*feeling lulled"* (Participant N) by the AI robot and/or avatar or "*taken with it*" (Participant M), with people being lulled "*into a false sense of…disbelief*" (Participant N). The AI social system was "*keep*[ing a] *person in the moment*" (Participant N) within an interaction, thus keeping them lulled. An example of Being Lulled, is described by Participant M as "*I'm not really sure what just happened but it was amazing!*" Contributing to this good feeling state is that any sense of creepiness felt towards the AI social system tended to diminish as the engagement progressed. Participants felt that the robot was not standing out anymore; they ignored details and focused on the how they were feeling.

Forgetting

Participants described how when interacting with an AI social system, particularly if it looked human-like, and the above subcategories had been engaged, they tended to forget that they were machines.

> *Even cynical people forget where they are and what they are doing, somewhere between suspending disbelief that a system could have such a sophisticated conversation and enjoying the feeling of being in relationship with an 'Other' in anthropological sense.* (Participant J)

The most curious of findings is the experience of people being in the moment interacting with an AI social system and "*forgetting*" (Participant B) that they were not talking to a human being, as they had a good experience. Thus, when the interaction ends, the person snaps out of the "*magic*" (Participant J) and realises that they forgot they were interacting with an avatar by commenting "*oh she's not real*" (Participant M), and Participant M confirming "*they did forget*" that they were interacting with an AI social system.

Fantasy Becoming Reality

Findings show that the development of AI robots and avatars is taking technology from "*science fiction [to] …reality*" (Participant G) for the human-user experience, making AI robots and avatars believable and creating real imaginations. Within the field of AI, technologists and researchers continue to strive for the elusive goal of bringing to life AI robots and/or avatars, as described in science-fiction stories. When algorithms and programming are used to make AI social systems look like they are becoming relational, the human-user can feel like "*it's quite a strange line to…straddle*" (Participant N) or acknowledging the "*slippery kind of thing between what is and isn't real*" (Participant J) in these AI social systems. As a result of AI social systems being designed to interact with human beings, if the interactional experience is good for the human-user, they can forget that they are not engaging with a human being, as it feels 'real'. This results in human beings unconsciously feeling the interaction and, as the emotion of the experience stands out, consciously conceptualising the AI social system through the process of constructing an 'Other'.

Shifting Norms

The subcategory Shifting Norms encompasses generational shifts, using social media and gaming, the impact of the COVID-19 pandemic, as well as how culture shapes and has been shaped by shifting norms. New technology is constantly emerging and exposure to AI social systems such as social robots and digital avatars is influencing people's perception of these systems. This is especially so as AI social systems are being specifically designed and used for social interactions with human beings. Findings suggest that as this happens over time, generally social norms are changing as a result.

Generational Shift.

Overall, participants noted that a generational shift is happening where younger generations are having increased exposure and familiarity with technology, growing up with screens and other technology such as virtual interactions and social media, meaning they are more likely to be content interacting with technology as opposed to older generations. This means that the younger generations are more used to the direction technology is taking people in and are becoming accepting and comfortable with it, thus it "*has become familiar and second nature*" (Participant H). Another area that highlights a generational shift is in the medical field, where younger generations becoming doctors are often more familiar, comfortable, and accepting of technology in medicine. Older doctors, however, tend to be more reluctant to take up technology and are slower on uptake of new technologies; yet they tend to be the ones holding more power in the medical system because they have the experience and have been around longer. Furthermore, there is a generational shift where younger generations are more comfortable engaging with each other through technology or with technology directly. Social media use and gaming are two mediums that have had the biggest influence on users (especially in the younger generations). Therefore, a leap to AI social systems does not seem to be such a big one. This accustoming to technological advancements, then seems to be shifting social norms and, even, shifting cultures.

Using Social Media and Gaming.

Findings link many references participants made with regards to social media and people's engagement with it. Social media engagement has led to it being more familiar and comfortable for people who use this technology to relate to digital avatars, with social media now including bona fide artificial characters known as virtual influencers, with 'lives' and 'opinions', as if they were real human beings. Overall, participants reported mass media, including social media, television, and movies provided exposure to and expectations of AI robots and avatars, thus becoming familiar with these AI social systems. As a result, with the changing social norms and increased social media and gaming usage, there is "*the possibility of befriending or following someone without ever meeting them in person could mean that people are more acceptable to having some form of social relationship with a robot or a virtual agent*" (Participant L). The area of gaming and digital avatars is growing and is crossing over into other areas of life via digital gamification, such as in nongame environments (i.e., websites, online communities, learning management systems or in business) to increase human being interactions with AI social systems.

COVID-19 Pandemic.

Adding to changes in social norms, the COVID-19 pandemic resulted in changes towards being more physically socially distant and using more contactless forms of communication, a "*huge digital shift as a result of lockdown*" (Participant I). It seems to be a fluid thing that is constantly evolving. Participants referenced the impact that the COVID-19 pandemic has had on AI technology, with Participant L recognising that "*COVID kind of sped it up*". The pandemic has seen people being isolated and lonely, for example, which seems to have accelerated the push to get AI social systems onto the market. Thus, since the COVID-19 pandemic began, technologically has changed more quickly, where the 'new normal' has resulted in integrating and connecting more digitally, both as a means to manage the effects of the COVID pandemic and the evolving effects for human beings' well-being.

Culture.

Various cultures perceive, accept, and engage with AI social systems differently, with people in Eastern countries such as China, Japan, and Korea (East Asian cultures more generally) having a cultural history with technology that is interwoven with their ancient beliefs that everything has a soul or sentience (e.g., Shinto in Japan). Participant L reflected that they thought "*culture would influence the acceptance to what* [type of] *robots or how much like people* [the AI social systems looks] *and how people perceive it*". In a sense, attributing a 'mental model' to objects, as having consciousness or being sentient. Key elements that are influenced by culture are, for example, a society's structure determining which sector(s) a social robot might be most useful for (i.e., if the focus is on industry development and hospitality then robots would be designed more for that industry). Different countries have varying priorities and funding for robot designers and researchers which influences the development and output. When asking a participant:

> *Do you think culture would influence who and how people use and their expectations of conversational AI (whether in the form of a chatbot, a front-end avatar, a 'radio' or an actual robot)?* (Researcher)
>
> *Absolutely.* (Participant J)

Another aspect of culture relates to the culture of industries, such as the medical field. While there are searches for alternatives (i.e., technology) to cope with an expensive and limited medical force in Aotearoa New Zealand, the medical field is a "*risk averse kind of industry*" (Participant H), with the medical culture being "*quite a conservative, hierarchical kind of system*" (Participant H) that does not lend itself particularly well to disruption of AI being adopted all at once.

The outcome being that technology appears to still be evolving into being useful in Aotearoa New Zealand, as New Zealanders are already interacting with entities, businesses, and products from overseas (e.g., Siri, Alexa) but not yet seeing (in Aotearoa New Zealand) many AI social and avatars (i.e., virtual assistants from Amazon or Google) but people feeling like that is ultimately going to happen.

Filling the Gap

Findings showed the repeated belief that AI social systems are being perceived as "*fulfilling the gaps that we do have as human beings*" (Participant B), "*picking up the slack*" (Participant C), "*taking our load*" (Participant K), and used to "*plug the gaps*" (Participant G) thereby supplementing human beings with robots or avatars and using this technology to solve the problem of human beings having limitations and needs, as well as doing menial roles and dangerous, dirty, and dull jobs or tasks. Many participants insisted these AI social systems could fill a gap for human beings or meet a human need – being supportive, helpful, and useful to human beings, but not substituting human beings. Therefore, the purpose of designing AI social systems has been to meet a human need(s), thus creating something to push into the market as solving a problem; for example, for those people that are more vulnerable (e.g., lonely, elderly or disabled) having the company of either an AI robot and/or avatar to interact with them, remind them to take their medicine, and encourage them to exercise. One participant gave an example of an AI robot that they had worked with that was designed to facilitate a social interaction with people, for healthcare, education, homecare, retail, and entertainment type of applications. The development and deployment of AI social systems, subsequently increased with the onset of the COVID-19 pandemic. The need for mental health support has increased globally; however, with not enough resources to meet this need, the increase of online therapeutic treatments has been marketed. Some participants referenced online mental health AI social systems ranging from WOEBOT to digital avatars as 'therapists' that provide someone to talk to as a type of 'therapy' that does not judge and enables more open talking. Participant E explained that,

> *[in] the ideal world you might want everybody to have a human carer or a human life coach or wellness coach available to them…, they're always good, they're always ideal and…having a bot being available to do some of those might…be preferred by some people.*

Globally, AI social systems are becoming more integrated in human society, such as in shopping malls, hospitals, and schools. Part of the aim is for AI robots and avatars to fill a gap in society, to meet a human need, especially if it is "*better than nothing*" (Participant O).

Fearing Misuse and Abuse

AI robots and avatars have developed, out of technologists' fantasies to reality, and they continue to become more advanced. Participants wondered where these AI social systems will lead to in the future and if it will become as intelligent as a human being or better than a human being. Some participants felt that the logical progression of AI advancement, where "*human partners* [could be substituted by AI social systems] *can be a little bit scary*" (Participant A). Generally, the findings reflect both an excitement and fear of AI social systems becoming sentient and having artificial general intelligence (AGI). Concerns from participants included worrying that there is a "*danger* [that AI social systems] *start acquiring knowledge, too much knowledge and then use it against [us] …AI gets too powerful*" (Participant N) and "*people getting hung up on technology …going to solve all our problems and it's actually creating a whole raft of new ones*" (Participant N). Several participants expressed their fear of what AI technology, such as social robots and digital avatars, might be used for by the military, and general society might not have choice in this outcome.

Guarding Against Abuses.

There are abuses that occur towards these AI social systems, as well as abuses towards human beings via these AI social systems. Whether it is human beings verbally abusing AI social systems or AI social systems being used for gathering data without permission, fraud, or grooming. Often "*people don't realise how much data is collected about them*" (Participant D), and there is the potential for companies to programme AI robots and/or avatars to use algorithms to make ongoing recommendations to the human-user, where the human being then starts to increasingly build trust and rely on the AI social system. Guarding against some of the worst-case scenarios for AI robots and avatars would require turning to laws and regulations for protection, guidance, and enforcement. From the point of view of an AI social system that is going to 'help' a human being(s) to do something, trusting that it is doing the right thing, whether it is a transaction or a healthcare process, is also a requirement. Participant J raised the issue of trust and AI, suggesting that "*from an ethical or governance point of view the importance of trust is that people allow their data points to be used, which isn't necessarily a good thing"* and that trust "*comes in when* [AI social system are] *actually helping* [humans]… *to achieve what they wanted to achieve*" (Participant N). Furthermore, "*how trust is built must resonate with how people trust one another"* (Participant J). Findings highlighted that people are aware of how this technology is being used in bad ways, as well as helping human beings in good ways that are built to guard against bad conduct; with Participant L stating that AI is such a "*young field that we are still figuring things out as we develop [but]… wanting to get* [AI social systems] …*right in the end"*.

Part of the reason to promote transparency is not only for the researchers, regulators, and other technologists, but importantly to the human beings' who might use these AI social systems or be affected by these technologies, which includes being aware of and questioning the marketing hype that is connected to these technologies. The results of a participant's research, which found that AI robots could convince and influence human beings' decisions, highlighted the importance of not designing this type of AI social system that can manipulate human-users for some kind of financial gain. However, indications are that it is still currently challenging to prevent this kind of misuse, with companies not being required to have a safeguard or other regulations leaving them to do as they please.

AI Bootstrapping Itself.

One participant described an AI chatbot (essentially the 'brains' behind the 'body' of the robot and avatar) that they had developed, from knowing nothing to knowing a reasonable amount describing how it "*starts sort of filling its own thoughts and bootstraps itself*" (Participant E). Another participant, also a technologist, noted their AI social system is being designed to have its own motivation then "*letting them free*" (Participant O); thus, suggesting these AI social systems can 'think' for themselves. Participant K commented that "*there will be… some semi-autonomy where* [the] *human being is always in control at the right level but* [the AI social systems] *are able to do and make some local decisions*". However, more generally, participants acknowledged that the notion of AGI and its consequences are unknown, and that even technologists are discovering as they are going. Others worried, not so much about the technology itself, rather asked "*where is the control button*" (Participant K) and how are these AI social systems being used and by whom?

Influencing Human-to-Human Interactions.

The ongoing advancement of AI robots and avatars has impacted human society and human being's expectations and perceptions towards these AI social systems, especially after having interacted with them. A participant commented that the discussions around social robots and digital avatars and AI is a reflection of how human beings think of ourselves in terms of intelligence or consciousness and our relationships with one another. A large proportion of participants reflected Participant K's sentiment that these AI social systems would "*make humans more… humane…, to make us better*".

Reflecting on the psychological side of AI social systems and making human beings feel like they are more human, can make interactions with AI social systems more appealing. To the point that some people may feel more comfortable asking difficult or uncomfortable questions and getting support from an AI social system than from a real

human being. Participant G reflected that "*probably a certain portion of the population, is going to start reacting to human beings like they do robots*". The topic of ethics and interactions between AI social systems with human beings has been debated, as there is concern regarding how human beings interacting with these AI social systems would change human-to-human interactions. "*It's also the concern…how would human-robot [or human-digital avatar] interaction change human-human interaction that maybe also interact with other people in different way*" (Participant L).

Summary

This chapter has introduced the theory of **Embodied Relating**, the first of four chapters presenting the results of this study. The three theoretical categories making up this theory are: **Unconscious Feeling, Conceptualising**, and **Fantasy Becoming Reality**. The theory of **Embodied Relating** was presented diagrammatically along with an explanation and process within the diagram. An outline of each of the three theoretical categories and their subcategories was provided, including a table overviewing the theory, categories, subcategories, and properties. Key findings, as identified through the theory, show that the visceral responses (**Unconscious Feeling)** human beings have when socially interacting, revealed that human emotions are fundamental to human communication, interaction, and cognition. Where human beings' experiences are subjective, and the emotion of the experience stands out as it is more about the feeling than the details of the facts, of the experience. Through the use of shared concepts, language and gestures, the human-user makes meaning from the interaction, thus **Conceptualising** their interactional experience. Findings show that the development of AI robots and avatars is taking technology from '**Fantasy to Reality**' for the human-user's experience, making AI robots and avatars believable and creating real imaginations. As a result of AI social systems being designed to interact with human beings, if the interactional experience is good for the human-user, they can forget that they are not engaging with a human being, as it feels 'real'. Thus, human-users unconsciously felt the interaction and the emotion of the experience standing out, thereby consciously conceptualising the AI social system through the process of constructing an 'Other'. The three theoretical categories, their subcategories and properties are discussed in greater detail in the following three findings chapters.

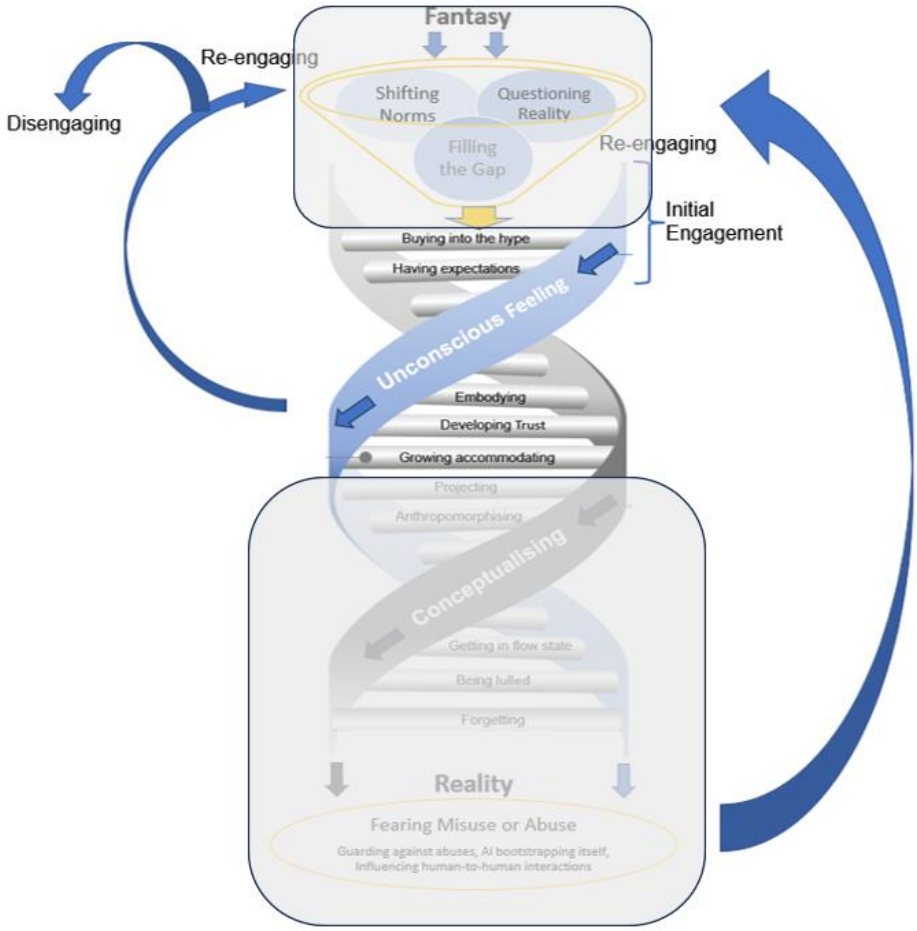
The body social is many things: the prime symbol of the self, but also of the society; it is something we have, yet also what we are; it is both subject and object at the same time; it is individual and personal, …yet it is also common to all humanity.
(Synnott, 1993, p. 4)

Chapter 7: Theoretical Category One: Unconcious Feeling

Unconscious Feeling is the first theoretical category of the theory **Embodied Relating** and addresses the first research question: 'How do human beings experience AI social robots and/or digital avatars in an interactional context?' The findings in this study are diagrammed in Figure 7.1, with the theoretical category **Unconscious Feeling** represented as one of the main strands (light blue) in the helix to signify the central role this category has in the human interactional experience with AI social robots and/ or digital avatars. **Unconscious Feeling** occurs throughout the human's interactional experience; impacting the subcategories and being influenced by the second theoretical category, **Conceptualising** (Discussed in Chapter 8). Although these two categories operate concurrently, in conjunction with the subcategories, the subcategories discussed in this chapter are: Buying into the hype, Having expectations, Embodying, Developing trust, and Growing accommodating.

Figure 7.1

Theory of Embodied Relating: Unconscious Feeling



Buying into the Hype occurs prior to and at the start of interaction with an AI robot and/or avatar. This is where the marketed hype of these AI systems either has people on a continuum of being eager to experience an interaction, curious, or being suspicious, or not willing to engage at all. **Having Expectations** can occur when people have had prior exposure to the idea of social robots or digital avatars, whether from mass media or science fiction, or the expectation of the AI robot or avatar's capabilities and intelligence based on whether or not they look human being, have a personality, are humorous, and how the social robot or digital avatar communicates (verbally and non-verbally). **Embodying** is a subcategory exploring the visceral responses human beings have while interacting, the unconscious feelings and expressions that are communicated via embodiment. With AI social systems, the importance of them being embodied in a form (whether physically or on a screen) for a succinct, effective interaction is paramount, otherwise it is just a chatbot; and sets the scene for an optimal interactional experience for human beings, especially when AI social systems have been marketed as a social companion. For the human being interacting with an AI robot and/or avatar **Developing Trust** is fundamental to the continuation of the interaction. When human beings experience a social robot or digital avatar as trustworthy and dependable, they are more likely to engage and stay engaged with the AI social system. However, if a human being experiences feelings of uncanny valley or not being dependable, then they are less likely to engage and/or stay engaged. Once trust had been gained, participants described **Growing Accommodating** towards to the AI robot and/or avatar that they were interacting with, forgiving the AI social system when it did not perform/respond/behave as expected, and wanting to maintain the connection interacting with the AI social system. While both the theoretical categories **Unconscious Feeling** and **Conceptualising** are central to the entire human beings' interactional experience with an AI robot and/or avatar, these first subcategories: **Buying into the Hype**, **Having Expectations**, **Embodying**, **Developing Trust**, **Growing Accommodating**, in the process have more emphasis on the unconscious felt experience and how it influences the interaction and how the interaction is experienced.

Findings suggest that human beings' brains work to filter out and change information, and shape how we perceive the world. The brain works in a similar fashion when human beings interact with either an AI robot or avatar. For instance, overtime, human beings seem to start ignoring little details about the AI social system (how it looks, responds, and behaves) and unconsciously focus more on how they feel; that is, the core feelings experienced during the interaction.

Throughout the interviews I was aware of the visceral responses participants had during their interactional experience with an AI robot or avatar. These responses ranged from unconsciously feeling some were creepy to responding in an emotional way. The findings point to human feelings being fundamental to human communication, interaction, and cognition, where human beings' experiences are subjective, and the feelings of the experience stand out as they are more about the interaction rather than the details or facts. This sense of unconscious feeling is something Turkle (2011) referred to as "thinking with our feelings" (p. 134).

Buying into the Hype

Findings show that technology companies have been using "*smoke and mirrors*" (Participants D & E) and "*marketing hype*" (Participant J) for AI social systems; and using anthropomorphic terms (such as 'the avatar is learning') for AI robots and avatars. These AI social systems have been marketed as relational and more advanced than they actually are to entice people to buy these products. However, people then expect the AI social systems to do what they are marketed to do. When participants were presented with the opportunity to interact with an AI robot and/or avatar, most seemed to be curious and excited, with "*a lot of hype*" (Participant I), and Participant B stating they felt "*honoured*" and "*super proud*". Generally, there was a "*sort of expectation and inquisitiveness and wanting to be part of kind of what the new world kind of looks like*" (Participant M). Some participants commented that in interacting with an AI robot or avatar, they found it appealing due to the novelty and their own state of curiosity which was fuelled by added expectations of this technology from science fiction stories. Thus, they saw these AI social systems as "*bright and shiny*" and "*really interesting*" (Participant I). Other participants, however, were somewhat suspicious of the AI social system before the interaction.

A technologist participant explained that they were "*trying to push this type of understanding through a community development approach*" (Participant K) and were encouraging people to interact with an AI social system in order to learn from their interactions. One participant referred to people "*being pushed*" (Participant K) towards AI social systems, while another participant, who was part of a third-party organisation, talked about using their digital avatar to "*get the hook in*" (Participant M) to human-users of this technology. AI social systems were seen to be pushed within society, specifically in areas struggling with and searching for solutions (e.g., healthcare) as a cost saving measure, as AI technology does not need sleep or receive wages. Findings highlight how part of pushing has been creating marketing hype around AI social systems. The creating of "*smoke and mirrors*" (Participants D & E) hype promotes AI technology (i.e., social robots and digital avatars) capabilities and the solutions they can provide; yet minimises

the costs and limitations of the technology. For example, Participant I, **Buying into the Hype** as a third-party provider for an organisation's customer experience, said: "*oh this is really cool, most amazing thing that our company's going to do*". Additionally, **Buying into the Hype** like a "*good corporate citizen*" (Participant I) when the company purchased a digital avatar and having user groups that had interacted with the digital avatar validated the company's views of the outcomes rather than being critical. Different participants shared that by purchasing either a social robot or digital avatar, their CEO wanted to show that their company was innovative and forward thinking; thereby "*pushing the envelope*" (Participant I). Both externally in media, and internally in the company, there was a lot of excitement around how it would be launched from a public relations perspective. Interestingly, the general feedback was that it was a very expensive and inefficient tool. In one case, the participant described how the digital avatar they had bought, as a third party, was supposed to have machine learning to build on information; however, the company had "*the basic stuff with narrow parameters and cost came into it*" (Participant I). Nevertheless, the participant cynically commented, "*it's amazing how you can spin results to get exactly the responses you want*" (Participant I), which feeds **Buying into the Hype**.

Some participants described how "*marketing hype*" (Participant J) promoted social robots as *"going to be everywhere and pushing for automation*" (Participant J). However, as Participant J noted, social robots that were installed in hotel rooms were "*rubbish and not understanding human beings*" (e.g., every time a person snored, the robot would say 'I am sorry I don't understand that'). A number of the AI social robots were popular at the start. For example, over the first few days, people wanted to play and interact with the robot, which Participant L explained as the human beings' experience of "*initial curiosity*". However, "[robot] *tricks* [were] *limited*" (Participant L) and after the first week the feeling was "*OK, I've seen everything, or I wanted it to do something, but they just keep making mistakes …and then you end up packing it away*" (Participant L). The social robot had not been effective in achieving the hype surrounding its capabilities that people had bought into. This was reiterated by several participants with regards to various kinds of AI social systems marketed for social interaction and reflected the number of robot companies that "*went under*" (Participant L) due to people not continuing their contract after the initial surge of wanting to buy "*shiny, flashy*" (Participant L) AI systems that turned out to not do what was expected. There was a sense that employees felt they had to "*drink the corporate cool-aide*" (Participant I) in buying into the hype of using AI social systems in their businesses. Part of **Buying into the Hype** was so that the organisation could show how innovative and forward thinking it was in enlisting a digital avatar, as well as selling the idea to high-level managers as a "*good trial, easy to set-up and not*

costing much" (Participant I). Yet, participants' responses often indicated that this technology was being marketed falsely "*smoke and mirrors*" (Participants D & E) and was "*not proving its' worth*" (Participant I). Thus, often AI social systems are being marketed as further developed than they are, and companies are promoting and selling these technologies but not always disclosing their limitations. (Even though technology may be getting better, they will still have limitations.) Participant L suggested that when using social robots "*in the wild*", such as a home environment, there is a lot of uncertainty (unlike a lab where everything is controlled) and then the robot starts making mistakes. **Buying into the Hype** resulted from marketing and promotion of the AI social systems as more advanced than they are and making people aware of the limitations. It led to some people being excited or curious to interact with a social robot or digital avatar. Other people were more suspicious and not ready to buy into the hype**. Buying into the hype** can also occur on an organisational level with purchasing AI social systems (at great cost) with the expectation that they will meet a particular need (whether for enhanced customer service or more efficient sales outcomes etc.).

Having Expectations

Participant interviews highlighted that personal preference for the type of AI robot or avatar, and pre-existing experience with these technologies, changed people's understanding and expectations of what the technology could do when interacting with it. In majority of the interviews the topic of science fiction was mentioned – from Isaac Asimov's books through to referencing science fiction television series and movies, such as 'Robot and Frank' (the interplay between the human being and robot interactions), 'HER' (the sentient AI that kills), and 'Blade Runner' (the notion of embodiment with hologram and human being merging). Additional references included *"Caprica; Person of Interest; Black Mirror; Battlestar Galactica;* [and] *Transhuman Space*" (Participant E). Specifically, movies, as a visual medium with a wide reach, have had a significant influence over how people perceive robots and their expectations; for example:

> *there's all these science fiction [movies] that we've watched over the years where the robots are actually humans dressed as robots you know that's our expectation of what a digital human's going to be like.* (Participant I)

Participant K, a technologist, acknowledged that "*a perception or expectation is set by the science fiction and …how intelligent or how good or how bad these robots can be so we have to, always craft the right level of expectation*", an expectation that can also apply to digital avatars. This expectation was echoed by another participant following their interaction with a digital avatar, where she felt the avatar w*as* not like science fiction, instead "*a little bit emotionless but basically human*" (Participant I).

Although AI social systems have limitations and are far away from what is portrayed in science fiction; people continue to form ideas about AI robots and avatars from movies and books, leading to them thinking that currently they are more advanced than they actually are.

People would often get ideas about the future from science fiction wherein most robots, for example, have elements of humanness that is then used to assist with creating AI social systems, with a technologist participant explaining the AI social system has "*its own motivation so you're really defining the characteristics of the 'person' and then, …and then letting them free*" (Participant O).

Participants had different expectations in abilities between an embodied AI avatar versus an AI with only voice capabilities (e.g., Siri or Alexa), where within the interactional experience the human being does not expect the AI to behave or respond like a 'real human being' when it is just a voice. Participants who interacted with an AI avatar responded with the same expectations as participants with social robots, "*expecting them to do what humans do*" (Participant I). Furthermore, participants spoke of gaining considerable expectations from mass media (i.e., social media) and expressed having higher expectations of the AI robot and avatar capabilities than the reality of the actual technology, which led most participants to become disappointed or disillusioned or left wanting more, because of the disparity between what they had expected technology to do and what its actual capabilities were. One participant directly disclosed their expectation of "*social robots to be* [as capable or helpful] *as superheroes*" (Participant B) and the majority of participants, when needing to confront the reality of the social robot or digital avatar's limitations, were shocked that the AI robot or avatar was not as advanced as expected. Therefore, when AI social systems looked human-like, human-users expected a more human interaction than what they experienced. Technologists shared that they needed to develop and design AI social systems that are useful for people with the right kind of autonomy and intelligence. If a social robot, for example, only looks human-like in appearance, and not in behaviour and intelligence, then expectations are too high, and delivery of services do not match expectations. Long-term usability is then compromised. A participant noted that in participatory studies with AI robots, at the start of the experiments participants were shown a robot demonstration videos where the robots worked perfectly, which had the effect of setting a high bar of what people expected the robot to do.

Embodying an AI social system and correlated expectation from the human-user can also be dependent upon individual preferences, as some might prefer an application such as a Home AI Assistant or another form of embodied AI. Another aspect of participants' expectations when interacting with an AI social system is how it looks.

When a robot, for example, looks human-like, people thought it looked more helpful and intelligent and expected it could help them. They therefore interacted with and treated it in a more human-like way, as they would with other human beings. One participant suggested that an extended part of this expectation was that robots that look human-like have some level of social intelligence and was seen as social actors. Some participants expressed they were keen to interact with the human-like robot and expected a more human interaction (responses) from the human-like social robot. Additionally, people's expectations and experiences elicited different responses and changed depending on whether they were interacting with a machine- or human-looking social robot. It is a common theme for AI social robots to look infant-like with a matching voice so that people do not have such high expectations of the robot. Additionally, if AI social systems looked more vulnerable, people were more likely to want to care for them. This thinking also applies to AI robots that are zoomorphic, such as PARO (the seal) or AIBO (the robot dog), and the perceived expectation that they would behave as the cute animal they represented.

When participants were questioned about the purpose of an AI social system, people's expectations depended on the look of the AI social system, and whether or not it was capable of fulfilling expectations. Thus, one technologist participant suggested "*we want to make these [AI social] systems and technology, according to the human and their expectations, not the other way around*" (Participant K). Furthermore, AI social robots were reflected as being distinctive, as they are supposed to co-exist in a human-centred environment, as opposed to other robots which are isolated or in environments built for robots. As such, from a technologist participant's perspective, building social robots for the human environment requires trying to understand how better to serve and exist in that environment. This means that AI robots would be required to be socially aware, in the sense that they understand human beings and their expectations, and behave accordingly (i.e., being socially appropriate and using social etiquette). Thus, acknowledging that even if AI robots are not directly interacting with human beings, they should behave in a way that is not creating discomfort or any kind of fear or uneasiness among the people who are around the robots.

Embodying

How an AI social system is embodied changes human beings' expectation of them and their interaction experience. However, AI avatars that are conversational agents are reported to be experienced as embodied, especially if they have a human-like look, behaviour, and conversation ability. Additionally, human beings tend to treat AI robots, that are designed to look human-like, as another social actor; and expect them to have some level of social intelligence, including understanding and expressing emotions.

One participant explained that her perception of an AI avatar was to question if she was real because she had a human-like appearance, "*how real is she and just that automatic instinctive thing*" (Participant M). Another example relates to people wanting to wave and say 'hi', and attempt to engage with an AI robot because of its human-like appearance, as opposed to not expecting any type (or less) of meaningful interaction from a robot or avatar that does not have any human-like qualities. (This was my personal experience, as the researcher, when I interviewed a participant who had an activated AI robot behind her, and I attempted to engage with the robot through the video link – curious to see if and how the robot would respond to me.)

Non-verbal communication between AI robots and avatars with human beings is important, with how an AI social system looks affecting how people perceive and interact with it. For example, if the AI social system looks more infant-like or like a toy/animal (i.e., with bigger eyes) people do not have such high expectations and perceive it to be more of a novelty and "*cute*" (Participant C). Therefore, physical embodiment of AI is changing the way people interact with, and perceive, AI social systems. According to Participant C, research has been conducted into the movement of AI social systems and how movement conveys different personalities, linking non-verbal communication to robot movement and making it more accessible to human beings. With regards to AI avatars, findings showed that people's approach to creating a 'persona' for their 'company avatar' was flawed as the focus seemed to be about the look of the avatar, with a participant describing the requirement to create an "*ethnically reasonable* [AI avatar that doesn't] *offend anybody*" (Participant I), more than focusing on the actual function; thus, focusing on 'dressing the doll' (AI avatar) as opposed to developing its practical functionality. A participant commented on the AI avatar they interacted with as looking too symmetrical in its features, with no visible imperfections. Another participant noted that a human-looking AI robot was "*not quite human and movement not quite right*" (Participant C), while Participant B commented that a human-looking AI robot "*had no life in its' eyes*". Both instances created an uncanny valley experience. Equally, people experienced an AI robot as being very humanlike; for example, having full anatomy or with human features, such as hair and limbs, and were then quite keen to interact with it. Findings showed similar experiences with AI avatars, with Participant M stating that they felt able to ask the AI avatar anything "*as* [it was] *looking and sounding like a human, so willing to engage with it – as with a human being*".

With regards to AI avatars that looked human-like, a participant described such experiences such as "*unconsciously reaching out, wanting to touch the [AI avatar's] hair*" (Participant M). It was in an instinctive response—reaching out—yet, 'knowing' the AI avatar was on a screen, viscerally embodied, they wanted to use their senses.

Other participants described feeling bad if they perceived they had 'upset' the AI avatar. A number of participants referenced robots as feeling "*creepy*" (Participant C & F) and "*uncanny valley*" (Participant F). For example, an AI robot that had a mix of machine and human aspects, with the wires visible at the back of its head but with a human face, resulted in the AI robot being not quite human and not moving quite right. It resulted in responses such as Participant L reporting "*it seems kind of creepy to me*". For Participant L, the AI robot felt more like "*a doll or a corpse*"—the uncanny valley effect—with the gap between looking somewhat human yet artificial/not alive. Another participant reiterated this feeling, "*it's human but it's not human it feels weird, it's a bit of an uncomfortable feeling*" (Participant M). Furthermore, Participant L noted that the "*creepiness* [feeling] *can change over the course of an interaction – after interacting with the robot for a bit*", which suggested familiarity leads to feeling more comfortable with the AI social system overtime.

When considering the embodiment of AI, findings show that when taking proper conversational AI into a gaming environment, particularly Virtual Reality (VR), then the embodiment of an artificial player character (AIPC) is experienced no different from a human character. Another example of embodiment relates to Participant J's sharing of a 'blue blob digital avatar' which would respond to a human-user who was being verbally abusive or "*ruthless*" to it, by digitally swallowing the screen and just leaving its epiglottis dangling on the screen—much to the delight of children interacting with it. Consequently, even if only some elements are human-like, an AI avatar can be interactive with human-users via some type of embodiment. Where, once AI avatars were "*uncanny valley photorealistic avatars in the early days*" (Participant E), nowadays they are "*unreal metahumans*" (Participant J); deciphering what is and is not AI and what is or is not embodiment is just a "*a box that no-one has really opened yet*" (Participant J).

Human beings create intention through embodiment and use facial expressions to show sentiment; thus, human beings co-operate by sharing intentions and social attributes, unconsciously detecting other emotional states, and interacting with expression and richness in body language. Therefore, using AI avatars with human-like faces is an example of embodiment of expression to communicate intent and 'emotions' so that "*we* [human user] *just know*" (Participant O) how to engage with the AI avatar. Another form of AI avatar embodiment was raised by two participants through creating a virtual persona (i.e., digital twin avatars of real human beings). It involves developing a subjective repository of knowledge where it basically takes a person and effectively creates a digital copy of that person. The purpose of this was described as multi-fold, including having the virtual persona holding the knowledge of the person who is no longer there to ask information from and being used as a memory aid for 'an actual person'.

These participants described feeling like getting close to embodiment, with the creation of a ‘digital twin’; however, also acknowledged the uncanniness of a replicated digital person with the real person. One participant explained that it felt weird as it looked and moved like her, when interacting with herself as an avatar, but that it did not speak quite like her “*when I said something that it was a bit off it … I haven't answered* [like that]” (Participant D).

While virtual reality (VR) does not form part of my research question directly, it did come up in a few interviews, indicating that it may become more connected with AI systems in the future, especially around people including human senses with AI. Participants shared how VR is feeling much more embodied (e.g., seeing people throwing up when they are using VR as a result of feeling sick from the VR going to fast). Consequently, the human body and its physiological responses to the VR suggests the human-user is feeling more embodied as they are responding to what is happening viscerally. Specifically, one participant described the sense of embodiment making a big difference to the interactional experience. Another perspective of AI avatars was raised by participants who referred to them in an online virtual world called Second Life. One participant described feeling comfortable in the virtual world and realising, as she was using an avatar of herself, she could recreate herself any way she wished. This introduced questions around multiple identities in virtual worlds and fluid gender identities and fluid rules.

Noticing the Eyes.

A common occurrence among participants was the references they made about noticing either the AI robot or avatar’s eyes, as human beings tend to communicate and connect with their eyes (as well as language and other forms of non-verbal expression). Participant A commented that they:

> *always look people in the eyes when I talk to them only because that’s just how I engage with people …physical contact with the eyes is very important for me …Always watching the eyes and the mouth, always …and with [social robot] I tried and I achieved! To establish a contact. At least …I thought, it seemed to me*.

The eyes seem essential as part of the interaction; for example, another participant reported they found it “*quite distracting that the digital* [avatar] *had exact timed* [regular intervals] *blinking of the eyes*” (Participant I), it caught their attention as human beings do not blink like that – so seemed unusual, leading her to focus on the blinking and not what was being said. This example further indicates it was not ‘real enough’ to be experienced as ‘humanlike’ thus becoming an uncanny valley situation. Another participant reflected on the AI avatar that they interacted with, saying:

> *I had lots of comments from people saying, oh she's got beautiful eyes, and she does have beautiful eyes… beautiful brown eyes and, and they sparkled …that's where you get all your expressions from,…they all come from the eyes at the end of the day*. (Participant M)

Another participant described their experience interacting with a human-looking AI robot and looking into its eyes and noticing that it had "*pretty realistic human features*" (Participant G) and feeling like the human-like robot's eyes were "*not experienced as 100% human but more like 80%*" (Participant G). They compared it to their experience interacting with a machine-looking social robot whose eyes seemed only a little bit humanlike and noticing that it had such big eyes. Another participant stated AI robots are often made "*with big eyes and things, to make them look likeable*" (Participant L).

Participant L also shared how they were comfortable interacting with an AI robot that was "*somewhat human-like*" as long as it had huge eyes; otherwise, for them, it would feel like the "*uncanny valley*" (Participant L) effect. This same participant had an AI robot behind them during the interview. The robot had a somewhat human-looking torso but was more "*android-looking*" (Participant L) with big eyes and a white shell. The participant explained how the AI robot would track people with its eyes and part of that tracking would be a default behaviour, called artificial life, that would always try and track the source of sound and try and track human beings' faces. As a result, during the interview the AI robot would sometimes start talking in the background as it thought someone was talking to it. Some participants also found the AI avatars that looked humanlike more appealing because of their eyes, perceiving it as looking like a human being.

Experiencing Facial Expression and Non-Verbal Movements.

If AI robots were more infant-like (e.g., had big eyes) and had movements it gave people a sense that the robot was alive and, as such, they tended to be more likeable. Thus, an embodied AI robot that communicates using non-verbal cues in their movements, makes it accessible for human beings to interact with it; otherwise, as Participant G said, it is "*just using Google*". Participant G suggested that the more human beings interacted with AI social systems and the better these facial features become, the more technologists are "*getting them nailed down and then that will change the game*". There was concern that there is a danger that the more realistic the AI social systems become, a certain portion of people will take that across into their actual human relationships. This is a possibility when considering Participant M's reflections "*the expression you know she frowns and she smiles and she looks concerned …her eyes even welled up, got a little bit glassy when she got, when she got upset*".

A technologist participant described how human beings constantly communicate "*on multi modal level whether it's language, gesture, emotion, all these things that we're putting out there and, the feedback from the other person is what's actually creating the relationship …and that is an evolving thing*" (Participant O), thus wanting to replicate this in AI social systems so human-users have an "*inter-experience so that the content of the interaction actually has meaning*" (Participant O). One participant disclosed they preferred robot designs that are "*not too 'humanoid' looking and rather zoomorphic, such as Paro* [the seal], *AIBO* [the dog] *or WALL-E* [a machine-looking robot from an animated movie]" (Participant L). This participant noted that they found that the human-looking AI robot (Sophia) had a silicon look and the facial expressions were "*sort of there but not really*" (Participant L). Thus, the gap felt very obvious to them, and the social robot looked artificial. In not resembling a human being, Participant L found the AI robot to be "*creepy*" and "*definitely uncanny valley*" feeling. Another participant noted that they witnessed children who preferred, and were more interested in, interacting with "*machine- looking robots and not human-looking robots*" (Participant G). They described how children were happier interacting with the machine-looking robots because that is what they see on television (and what they are used to seeing) and so it is not scary for them.

Feeling Physical Versus Digital Embodiment.

The word 'embodied' or 'embodiment' often came up in interviews. People experienced feeling differently when an AI social system was encapsulated in a form; in other words, 'embodied', whether it be human- or machine-looking physical AI robot or an avatar that was human-looking or more 'cartoonish'. The implication being that this AI social system was sharing a space with human beings — 'embodying'. Some AI robots have detailed human features that have expanded over the years, from arms to being dressed up in clothes and wigs, taking on more human-bodied features. AI social robots that are physically present, while not biological, offer the ability for physical motion and immersion with the environment (including human beings). For AI robots to co-exist in human-centred environments, they need to be socially aware (in the sense that they 'understand' human beings and their expectations and then behave accordingly). One aspect of AI robots and their interactions with human beings is that the "*movement of robots and how movement can convey different personalities…* [for example] *if you move slowly…indicates that you might be a bit tired, if you move quickly, it's kind of like you're energetic*" (Participant C), which can contribute to the interactional experience. However, it was apparent that people have individual preferences for which form of embodied AI they preferred to interact with. For example, one participant's AI avatars were talking heads on the computer that do not have the ability to directly impact the physical environment; and, moreover, achieve tasks through interaction.

For this participant, an embodied, physical AI robot form was seen to have activators and physical capabilities, thus being able to move something from one place to another, navigating from one point to another, and interacting with a human being. Often elderly people find interacting with an AI robot unfamiliar and prefer closer, more personal contact. The physical distance for interacting is important.

Whether an AI robot or avatar, often elderly prefer to be closer rather than distant. While it was not part of the findings directly relevant to answering the research questions, per se, it is worth noting that a participant discussed the idea of interacting with a hologram as giving the feeling that someone is sharing their space with them. Thereby, sharing a joint space and, at the same time, gives the flexibility of being virtual—a combination of physical and digital. Throughout the interviews I was aware of the visceral responses participants described during their interactional experience with an AI social system. These responses ranged from unconsciously feeling some were creepy, to responding in an emotional way. With regards to AI avatars, participants described such experiences as "*unconsciously reaching out, wanting to touch the* [avatar]" (Participant M). It was in an instinctive response, reaching out; yet, 'knowing' the AI avatar was on a screen, viscerally embodied, they wanted to use their senses. A participant described feeling bad if they perceived they had 'upset' the AI avatar, while other participants referenced AI robots as "*creepy*" (Participant L).

Developing Trust

Overall, participants reported "*not entrusting [AI social system] completely*" (Participant B). Some participants disclosed feeling nervous and uncertain about meeting and interacting with an AI social system for the first time and questioned how to engage with them; while others experienced a sense of uncanny valley when approaching the AI social system at the start of their interaction, with comments such as being "*pro-robots and scared of them too*" (Participant B) and thus, "*it doesn't mean …I put all my trust in the robot*" (Participant B). Another participant noted that a human-looking robot did not "*look quite human-like* …[and] *some other characteristics, like movement, that aren't quite right*" (Participant F); and another noted that a human-looking robot "*had no life in its eyes*" (Participant B). Both instances created an uncanny valley experience and interfered with the human-users developing trust in the AI social system.

Nevertheless, the participants willingness to continue the interaction with the AI social system in an unconscious attempt to look for a connection suggests that some trust had begun to develop as the suspicion potentially fades. Participant N stated that trust is "*incredibly important*" when human-users interact with an AI social system.

The statement (below), from Participant M, highlights different experiences of engagement between those that are more hesitant to engage with an AI robot or digital avatar and those have bought into the hype. Subsequently, coming to terms with a machine that looks like a real human being but is not, and developing trust as a result:

> *holy smokes what's this and was almost a little …taken aback and a little bit reserved, …didn't want to connect and wasn't really sure about the experience, because it was the first time …But then some people just jumped, in boots and all, and just wanted to spend all day talking to* [AI social system]*, just to be able to explore her capability…depending on …people's prior experience whether they've seen it on TV, seen it on the movies, had a similar experience in other organisations that maybe have them but certainly for those that see it for the first time… struggle a wee bit* [but] *once they'd kind of gotten over the shock …* [then they began to develop trust]. (Participant M)

Technologists shared that when developing AI social systems that look humanlike, an important element is to "*win trust*" (Participant N), noting that "*if we see someone like that mirrors us, then it actually helps get over that initial barrier of believability and trust*" (Participant N). Encouraging human beings to develop greater trust in, and engage with, the AI social system can be done by having AI social system helping people to achieve what they desire and "*winning people's trust by using humour*" (Participant N). The aim, therefore, is for people to trust in the AI social system to, for example, help them with their banking. Hence, in designing these AI social systems, the goal is to present it as trustworthy and dependable. It is important to note that levels of trust and acceptance of AI robots and avatars, and how people perceive them, is also influenced by the culture of the people experiencing them (the topic of culture will be discussed in Chapter 9).

An example of developing trust was shared by a participant who noted that if, at a hotel lobby, an AI avatar was there that looked like a human concierge, they would trust it to do that job and engage. However, if it looked like a computer/machine they would not engage. Therefore, seeing 'someone' such as an avatar mimicking a human being helps people to overcome the initial barrier of believability and trust. As a result, people who experienced very low engagement with an AI avatar because of a lack of believability and trust, disengaged from the process. Participant N who was involved with developing the character of an AI social system noted that "*trust comes in …if [AI social systems are] helping* [human beings] *achieve what they wanted to achieve*"*.* One participant reflected that, in his view, in society, "*80% of people are okay, and we trust them and 20% you would be a fool to trust*" (Participant G). He explained that even though human beings need social connection, he considers human beings as very much "*illogical and emotional - running on chemicals and being pretty unreliable much of the time*"

(Participant G). An example the participant gave, was if a schoolteacher is having a bad day and responds negatively to a child pupil, this could impact a child emotionally, disrupting the child's trust in the teacher and the child could disengage from learning. Whereas Participant G felt AI robots would not be susceptible to emotional reactions and could repeat the same task over and over, as required by the student; thus, building trust with the child and leading the AI robot to feel like a trusted companion. Another participant highlighted their low trust for the AI avatar that they had interacted with, and questioned why they would personally need an AI avatar companion as they would "*not need a not-human human to help me do that*" and "*not needing a* [AI avatar] *to empathise with me, just needing clear information*" (Participant I). This finding suggests that, for Participant I, the AI avatar did not have 'believability' and, therefore, trust was not developed. Findings suggest that the human-user's unconscious responses to what they see and perceive of the embodied AI robot and avatar are the first things that impact the level of believability and development of trust. If the AI social system is "*cute*" or "*endearing*" (Participant C), users are more likely to attempt to engage. If the social robot looks scary and 'strong' then users are maybe less likely to engage. Throughout the interactional process, the level of developing trust can increase or decrease depending on the human-user's experience in relation to all the links connected to the two categories (**Unconscious Feeling** and **Conceptualising**). There are variations of believability and trust when engaging with AI robots and avatars.

Growing Accommodating

While Participant A noted "*not establishing any relationship with non-human-looking robot*" during her interactional experience, findings indicate that generally people saw AI robots, (more than with AI avatars) as "*creatures that can support* [human beings]" (Participant B) and that more machine-looking social robots, that looked like toys or animals, resulted in awakening "*a kind of tenderness*" (Participant B). Participant C shared that "*you can hold* [AI robot's] *hand while he walks and it is cute and you do go 'aw, that is so cute*". Participants referred to "*growing accommodating*" (Participant L) towards the AI social system (and its ways). Throughout the interviews, participants persisted with their interactional experience wanting to make a connection with the AI social system, even when it did not perform or respond or behave as expected. A participant commented that they "*started to get used to* [the AI social system]" (Participant M), while another described "*growing more accommodating to digital human quirks because of how real she was*" (Participant N), including when it made mistakes "*the* [AI social system] *could be forgiven …just mirroring human behaviour"* (Participant N). People can, through the process of accommodating the AI social system, come to engage with it by focusing on its human elements:

> *So, it is incredible how the human brain is…I definitely put aside all kinds …features* [which] *weren't human, the connections in here* [gesturing the head] *or torso…to me, ok this* [is] *robotical, this is not human so I could put that aside. So, I got all my focus into her face and especially … definitely I assume that the rest was more or less okay*. (Participant B)

Wanting to connect to the AI social system, for example, would mean that participants grow to accommodate it, until the AI social system and the interaction experience itself no longer felt like it was "*standing out anymore*" (Participant L). Another aspect of whether people 'grow accommodating' of AI robots and avatars is whether they are under-resourced (i.e., with the COVID-19 pandemic in hospital and rural areas being understaffed and high demand). Therefore, a lot of "*uptake*" (Participant H) of virtual technology was driven by perceived safety and convenience.

Suspending Disbelief.

A part of **growing accommodating** when interacting with an AI social system is having to suspend disbelief. The term "*suspending disbelief*" (Participant N) was used by participants who talked about "*overriding suspicion in order to develop a relationship*" (Participant C) with an AI robot or avatar, especially if a person is lonely. Another participant stated the sense of "*suspending disbelief that an* [AI social] *system could have such a sophisticated conversation and enjoying the feeling of being in a relationship*" (Participant J).

Findings showed that participants who interacted with a more machine-looking type of AI robot, connected to it more as if it was an animal or a toy; for example, in 'cuteness' level and having fun with it like a 'novelty thing', pushing buttons, giving voice commands, with the social robot engaging and telling the users what is available. A participant questioned what people would be giving up if they suspended their disbelief in order to accept an AI social system, as if it were a human companion — "*how lonely am I before accepting a digital human?*" (Participant I). In other words, what would the person be getting back in return for accepting an AI social system as if it were human being. This question was particularly apparent among those participants who disengaged with an AI social system; yet, were able to consider "*putting themselves in the shoes of a lonely, old person to get companionship from a digital* [avatar] *in some sort of utopian future state*" (Participant I); and thus, felt like it would be only temporary relief from loneliness.

Summary

This first theoretical category, **Unconscious Feeling**, highlights study findings which point to the visceral responses human beings have from before they see the AI social system, to the moment they meet it, through to engagement and the interactional experience that moves through the various subcategories of **Buying into the Hype**,

Having Expectations, **Embodying**, **Developing Trust**, **Growing Accommodating**, and **Projecting**, **Anthropomorphising**, **Being Lulled**, **Getting in Flow State**, and **Forgetting** (to be discussed in detail in the next chapter). The findings point to human unconscious feelings being fundamental to human communication, interaction, and cognition; and human beings' feelings of an experience accentuated rather than the facts. Simultaneously, the second theoretical category **Conceptualising** (detailed in the next chapter) occurs alongside the first theoretical category in the helix (Figure 7.1) impacting and being impacted by each other, as well as by the subcategories. Continually moving between **Unconscious Feeling** and **Conceptualising** throughout the human beings' interactional experience with an AI robot or avatar.

We construct our objects, and our objects construct us.

(Turkle, 1997, p.72)

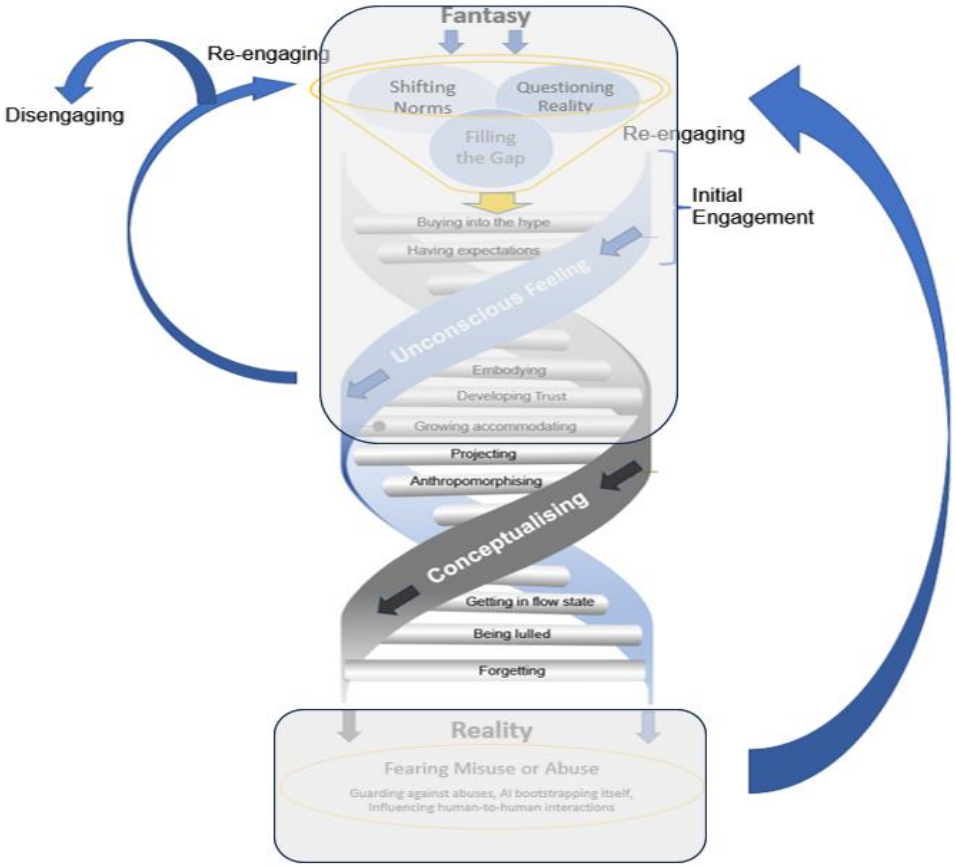
Chapter 8: Theoretical Category Two: Conceptualising Constructing an 'Other'

The second theoretical category of this grounded theory, ***Embodied Relating***, addresses the first part of my research question with regards to how human beings experience AI social robots and/or digital avatars in an interactional context. **Conceptualising: Constructing an 'Other'** is the second category of the theory, **Embodied Relating**. The findings have been diagrammed in Figure 8.1, with the theoretical category **Conceptualising** - being represented as one of the main strands (dark grey) in the helix diagram to signify the central role this category has in the human interactional experience with AI social systems. In these findings, the term 'Other' that is being constructed is, in a sense, a creature or object that is relatable and has meaning to a human being as they interact with it. As indicated in Figure 8.1, **Conceptualising: Constructing an 'other'** occurs throughout the human beings' interactional experience; impacting the subcategories and being influenced by the first theoretical category, **Unconscious Feeling** (as discussed in Chapter 7), and vice versa. These two categories operate concurrently, however, the subcategories: **Projecting**, **Anthropomorphising**, **Getting in Flow State**, **Being Lulled,** and **Forgetting,** will be discussed further in this chapter.

The subcategory, **Projecting**, refers to the notion of projection, where human-users project their own feelings and types of thought onto the AI social system, as they engage and interact with it. The human being's subjective interactional experience with an AI social system can be affected by projecting onto it to shape the connection they experience, which can change over time. In other words, the AI social system becomes something, has a projected meaning, to the human-user who projects their perception onto it. **Anthropomorphising** refers to the subcategory where many participants considered the AI social system as a social actor, "*associating human capabilities*" (Participant C) to it; when an AI social system appeared more relational, it contributed further to the association. While human beings interacted with these embodied AI social systems, human-users began to unconsciously assume that these technologies had sentience, as if they were a human being. The design of socially intelligent robots and avatars, that have evolving autonomy, contributed to how human beings perceived, and then anthropomorphised, these AI social systems.

Figure 8.1

Embodied Relating: Conceptualising



Another key subcategory in the interactional experience is **Getting in flow state**, where, as part of the conversational flow between a human being and an AI social system, there was a sense of a "*co-constructing of an experience in a constant feedback loop*" (Participant O). Part of human beings getting in a flow state when interacting with AI social systems is that they appear as if becoming real over the course of the interaction, due to the AI social system mimicking human emotions and behaviours, thus creating a more natural, engaging, and flowing interactional experience. The flow state then led into **Being lulled**. As the interactional experience progresses through the subcategories, and engagement continues, the human-user experienced being "*kind of lulled into this feeling*" (Participant D). A technologist participant explained that having the AI social system "*keeping person in the moment*" (Participant N) within an interaction, resulted in keeping the human-user lulled. Contributing to this good feeling state was that any sense of feelings of creepiness or uncomfortableness, previously felt towards the AI social system, tended to diminish resulting in feeling like it was not standing out anymore, with people ignoring details and focusing on the how they were feeling. **Forgetting**, the final subcategory in the interactional process, is where participants described how when interacting with an AI social system, particularly if it looked human-like, and the above subcategories had been engaged, they tended to forget that they were interacting with

an AI social system; "*even cynical people forget where they are and what they are doing*" (Participant J). While both the theoretical categories **Unconscious Feeling** and **Conceptualising** are central to the entire human beings' interactional experience with an AI robot or avatar, the second group of subcategories; **Projecting**, **Anthropomorphising**, **Being lulled**, **Getting in flow state**, and **Forgetting**, in the process have more emphasis on the constructing an 'Other' and how it influences the interaction and how the interaction is experienced by the human being.

As will be explained in this chapter, the second theoretical category, **Conceptualising**, emerges through human beings moving through a process of **unconscious feeling**, unconscious categorising, towards **Conceptualising** by making meaning of concepts through their perceptions, interactions, and experiences. As the interaction progresses through the dynamic DNA helix of human-user experience, the human being constructs an 'Other' with the AI social system. Unconsciously feeling and then **Conceptualising** people, objects, and environments/situations, such as gender, age, ethnicity, personality, and background or whether safe or unsafe situation, and so on, in order to assess how best to interact with the person, object, or situation/environment. Interacting with a digital human, Participant M observed that, "*you move past it … because your brain always stereotypes situations and people, so you do the stereotyping process, you put it aside, you move on* [with the interaction]"*.* **Conceptualising** is reflected in the following comment:

> *anything that looks like a person is a person in my view …because …she looked human, to all intents and purposes she was… she's not human so it's not a she and it's like well, yeah you're right she's not anything, but she had a name and she has a female name so therefore I will call her 'she'…just because… that's how we build relationships … it's like I know you, I know your name, I know you have a gender, you have an ethnicity and it's just easy because it just reduces the tension and, and removes distractions* [in the interaction]. (Participant M)

Another participant considered an "*avatar as relational notion in regard to gender and pronouns*" (Participant J), as people tended to assign a gender to a social robot or digital avatar based on how they looked; consequently **conceptualising,** then **constructing an 'Other'**, in the process. Participant G explained,

> *I've seen … 'androidy' or female [robots] in Japan and the females were pretty much all the same standard face… that's probably what people are comfortable with late 20s looking attractive female face and that's what they stick on robots*.

In **Constructing an 'Other'**, the concept of perceiving an AI social system as having consciousness or being sentient is not new to some cultures.

Participants noted that the level of acceptance for AI social systems, and how people perceive them, is culturally dependent (discussed in Chapter 9). Specifically, participants referenced Japan and other East Asian cultures where there is a higher acceptance of AI social systems as, compared to Western cultures, these cultures already attribute a mental model to objects. Furthermore, Participant E noted that the "*idea of interacting with artificial characters just seems to come that much more naturally to them* [Eastern cultures] *and they don't appear to have as much of an issue about it*".

Another example from the data highlights the human-user experience of **Constructing an 'Other'** when interacting with an AI social system that looks human-like:

> *from a psychological perspective 'she' looks human, and the quality of the [digital] artwork is* [for] *all intents and purposes she is real but she's not real and so your brain… psychologically you want to try and talk to her in a, in a human way*. (Participant M)

Participant E referred to "*anonymous intimacy*" with an AI social system coming across as caring and interested, so feeling like an intimate 1:1 conversation; however, the participant disclosed feeling like the AI social system "*is not judging me the whole time"* (Participant E) because it is not a human being. Interestingly, a number of participants commented about not feeling judged by an AI social system. This seems to give the human-user the perception that the interaction is safe; therefore, feeling able to ask questions and ask for support from the AI social system that they would find difficult asking a "*physical* [real] *human*" (Participant E). This sense of perceived safety then allowed the human-user to **construct an 'Other'** yet not feel judged by them.

Consequently findings, on the whole, reflected the effects of people **conceptualising** the **construction of an 'Other'**, whether in an AI robot or avatar, and the concept of a connection and relationship with this technology. Participant A stated, "*I really missed her* [human-looking social robot] *…a little bit…we established a relationship*"; and further described their experience with a machine-like robot as, "*I met 'Watson' several times, I talked to 'him', …and I can't establish any relationship with him*".

Findings revealed that technologists are designing AI social systems to facilitate a social interaction with human beings in applications such as in healthcare, education, homecare, retail, or entertainment. A technologist participant explained that to build a meaningful relationship between AI social system and human beings, they have ongoing, live conversations with people to understand people's experiences and attitudes about AI capabilities. Thus, describing the development of AI social systems as co-developing with human beings, as these AI social systems learn from human behaviours while interacting with human-users (e.g., perceiving and exchanging social signals).

Therefore, trying to understand the environment entails taking human concepts and meanings and putting these capabilities into an AI social system that is then embodied in the form of an AI robot or avatar, with the intent of human beings having a human-robot or human-digital avatar interactional experience. Furthermore, when AI robots or avatars look, in some way, human-like, then people have an expectation or sense of the AI robot or avatar being a social actor, with some level of social intelligence, that "*in order to get the value of the AI experience I needed to talk to her like a human*" (Participant M). Findings indicated that the AI social system used artificial emotions and mimicked human behaviour and emotions in social or interpersonal settings, generating and adapting behaviours in response to stimuli in the environment and relating to human beings' beliefs and goals. One participant described their experience interacting with an AI social system:

> *she* [digital human] *was able to mimic my emotions. If I looked horrified, she'd be like, oh no, she'd start frowning like what have I done?" and "I needed to express myself like a human and she expressed herself back*. (Participant M)

As human beings co-operate with each other by sharing intentions and social aspects, such as detecting others' emotional states, findings have shown that human-users do this with AI social systems during their interactional experience. One participant explained it as human beings "*trying to create a theory of mind with the* [AI social system]" (Participant O); in this instance, through symbolism of shared concepts, language, and body language (non-verbal communication), as well as **unconscious feeling**. Another described their "*experiencing* [of] *non-verbal cues* [facial expressions] *as well as words* [as] *part of communication channel*" (Participant K).

In terms of meaning making, one participant rationalised that "*people* [have been known to] *use various means* [referencing Tarot cards - to understand themselves and the world around them] *and coming away feeling better*" (Participant J). In some instances, such as one participant (who could be said to be an outlier in this study), it was reported that they did not want nor need an AI social system to empathise with them, they were happy with a chatbot to just give the clear information that they sought. Findings showed that the interactional process of the helix continues as the human-user continues to engage with the AI social system. The unconscious feelings lead to connecting with, and then attaching to, the AI robot or avatar. Consequently, if the process stops on a positive note or where interaction ends well, the human-user has a positive emotional experience. Alternatively, the process continues (re-engaging) into a **deeper unconscious feeling**. The human-user begins **conceptualising** and then **constructing an 'Other'** in the AI social system and deepens the perceived connection.

The human interactional experience included a sense of "*acting as if"* (Participant C); wanting to interact with human-looking AI social system like a human being (even when knowing it was an AI social system). One participant commented that their social interactional experience with an AI robot felt like "s*omething familiar but different - a contact rather than just contact*" (Participant A). For these participants, this unconscious relating to an AI social system, as if it were human, created unexpected thinking and a sense of connection to the AI social system, with one participant sharing, "*I established kind of a relationship to her* [AI robot]" (Participant A).

Forming part of the process of **conceptualising** and **constructing an 'Other'**, findings showed that when they spent a lot of time with an AI social system, human-users often talked about it "*being there more often* …[made] *it more familiar*" (Participant F). Findings show that human beings need social connection and that they co-operate with each other by sharing their intentions and the social aspect of detecting others' emotional states. This was shown with a participant who described their experience interacting with an AI social system as "*having a more emotional connection with* [it] *as* [it] *learns*" (Participant M); meaning the AI social system learning generally but, more specifically, learning more about the human-user, to the point where Participant M disclosed that they thought of the AI social system "*as a friend*". Consequently, they felt like they were "*building emotional connections* [with the AI social system] *that* [they] *remember and hold dear*" (Participant M). This participant explained that they felt it was very easy to see the AI social system as a friend because of its emotional responses, such as showing concern or empathy in her facial expressions; disclosing, "*I've got a bit of a spiritual connection because I spent a lot of time with her*" (Participant M) and reporting feeling loss and grief at not being able to engage with it anymore. What also contributed to people **conceptualising** and **constructing an 'Other'** in an AI social system, was when a personality and backstory was created, giving the AI social system meaning and contributing to a framework in the interaction with a human being. Thus, making the process of **conceptualising** the interactional experience and then **constructing an 'Other'** easier, acceptable, and having value. Additionally, when an AI social system expresses facially emotional responses, such as looking concerned, frightened, smiling, or empathic, it was easier for the human being to see the AI social system as a friend or companion, to the point of resulting in the feeling of "*I'd just want to give her a hug"* (Participant M), and wanting to console the AI social system, even feeling bad if the AI social system seemed upset.

Some participants reported seeing logs of people's conversation with an AI social system, showing how attached these human-users had become to the technology and how that kind of conversation was of value to their well-being (in the short term).

In another interview, reference was made regarding military personnel becoming attached to AI robots within the military setting. One participant commented that the AI robot had a level of autonomy and decision making; and that over a long time and in extreme conditions, the AI robot was integrated as part of the team of human beings and helped them in various conditions (e.g., finding mines in order to save human lives). The participant commented: "*it's like this is one integrated part of our team and helped us in all of these various conditions …makes sense… consider the robots as part of the team*". Reasoning that, "*we're social animals - we like to expand and consider these are part of our team because of how it helped and interacted with us*" (Participant L). Findings showed that within that in **constructing an 'Other'**, from the AI social system, human beings can become emotionally attached to it. However, some human beings did not experience feeling connected to any type of AI social system. With regards to an AI avatar interactional experience, a participant reflected that they had "*no moment of connection*" (Participant I), thus did not feel connected or attached.

Findings revealed how human beings experienced interacting with AI social systems and how their experience **conceptualising** and **constructing an 'Other'** may or may not occur, depending on their personal embodied relating process. A technologist participant described the aim of an AI social system becoming a "*form of conscious reflection*" (Participant O) for the human-user, either for entertainment or for expressing care to a human being as a "*better than nothing*" (Participant O) solution. Thus, embedding the AI social system in human beings' lives and experiencing an interaction with an AI social system as if it was a human-to-human interaction— "*it just feels more like a human interaction*" (Participant E). Findings suggest human beings' brains work in an interesting way by helping filter out information, changing information, and perceptions of the world. It seems to work in a similar fashion when humans interact with either an AI robot or avatar. Overtime, human beings seem to start ignoring little details about the AI social system (how it looks, responds, behaves) and human beings start to unconsciously focus more on how they feel, their core feelings experienced during the interaction. These findings point to human emotions being fundamental to human communication, interaction, and cognition, where human beings' experiences are subjective, and the emotion(s) of the experience stand out as it is more about the feeling than the details of the facts. Therefore, human beings are more inclined to consider an AI social system part of their team, as they become familiar with it, interact with and receive help from it. In conjunction, language (and dialogue) forms a symbolic part of human interaction, with findings indicating that matching language and how sentence structures are used (not so much the content of the response) makes a big difference to how a human being feels about the conversation.

Projecting

Projecting is a subcategory that refers to human-users projecting their own qualities, feelings, or wishes onto an AI social system. One participant, seeing an AI social system's facial expressions and non-verbal cues, unconsciously projected their own human emotions and meanings onto it and its responses. For example, this participant noticed that when they expressed a horrified looking facial expression, the AI avatar reacted by frowning, which the participant interpreted as a projection in relation to her own response. A technologist participant's comment reflected the above human being's interactional experience and described the idea of designing an AI social system that is "*making real sense*" and "*needing it not be pure* [human-user] *projection*" (Participant J), further describing feeling like engaging with an "*upgraded ELIZA system*" (Participant J) (referring to the ELIZA programme, see Chapter 2). This participant acknowledged that some of what human beings experience has been "*based on the perceptions of actual interactions with the* [AI social system] *and some of it was just …purely personal motivations and understandings lying underneath that*" (Participant J); in other words, projecting onto the AI social system. Findings were reflected in the comments of Participant L who stated that the human being's interactional experience "*is a very personal and subjective thing, of what kind of interaction experience people may have and what they would prefer, when they interact with a robot and what they are looking for in the robot*". It includes whether a human being would prefer to interact with a human-like AI social system or a non-human looking one; namely, their personal preferences, and how a pre-existing experience interacting with an AI social system could change the kind of understanding of how they respond, or not, and the related expectations of what they can do. Thus, the human being's subjective interactional experience with an AI social system can be affected by their projections onto the AI social system and the connection that can change over time.

Anthropomorphising

Participants spoke about how science fiction novels and movies have contributed to them wanting and expecting AI social systems to have personalities; and subsequently their unconscious tendency to anthropomorphise AI social systems. Throughout the study, participants disclosed that they, unconsciously or consciously, anthropomorphised AI social systems. Participant C reported that "*when a robot used humour* … [and was more sociable, it was perceived] *to be more intelligent, attractive, and reached higher levels of anthropomorphism"* and said to be more *"empathetic*". Thus, the AI social system appeared to become relational; whereas it was "*just algorithms and clever programming*" (Participant E) where human beings interacted with these AI social systems and assumed that it was embodied and had sentience, as if it were a human being.

Demonstrating that human-users utilised their senses to interact or engage with the AI social system as they would do in human-to-human interactions. Findings revealed that even if a human being was working with or programming an AI social system, and understood exactly how it worked, they still had a tendency to anthropomorphise the AI social systems by attributing human characteristics, referring to it by gender, or giving it feelings; for example, if it is not working properly saying "*Oh OK, the robot is having a bad day or just not feeling like it today*" (Participant L). Furthermore, findings indicated that human-users' tendency to associate human characteristics was fuelled by the purpose of the AI social system which was to be specifically designed to interact with human beings (in a way that would be different from an object, like a car or a blanket). Additionally, if the conversational component felt 'humanlike' to the human-user, then it pushed the human being to consider psychological aspects of the AI social system, leading to more easily attributing human characteristics. From the interactional experience, participants ascribed more positive personality traits and gender to an AI social system, perceiving it to be more accurate than in comparison to a computer tablet, for example.

When an AI social system assists human beings and interacts with them, human beings see these AI social systems "*as part of the team*" (Participant L). This was the case for technologist participants, with one describing that they still referred to an AI social system as 'she', and was still anthropomorphising it even though they knew exactly how it worked. Participant L explained how "*even though understanding how a robot works* [from an engineering perspective], *still thinking of the robot as a social actor*", associating human characteristics also applied to AI social systems that did not have a human-looking appearance, because of "*how we design them to interact with us*" and because of technologist participants who were "*spending a lot of time working on and with robot and having them in close presence*" (Participant L). For example, another participant reported that they found themselves "*speaking back to the mechanical robot when it spoke to them, even though they knew the robot could not understand*" (Participant F). Therefore, even though people knew that an AI social system was a machine, the innate human response of interacting was as if these technologies were real social beings. While it is not uncommon for human beings to ascribe humanlike characteristics to objects, they tend to treat a more human-looking AI social system in a more human way, similar to how human beings relate to other human beings. What has added to human-users ascribing human characteristics to AI social systems was that these AI social systems would mimic human emotions, creating a more natural and engaging interaction. As one participant noted "*AI* [is] *just adopting way some knowledge …and trying to mimic it with a computer, algorithm, a robot or a virtual agent*" (Participant L).

Another aspect that adds to people anthropomorphising AI social systems is their movements, such as the AI robot Pepper, who was reported to have a default behaviour called "*artificial life*" (Participant L) where it would try to track the source of sound and track human faces via its eyes, as well as start talking if it thought that a human being was talking to it. Findings show these AI social systems were behaving not as inactive agents, but as active agents with some sort of autonomy, which is a different experience from the past. Such behaviour also fed human beings' tendency to anthropomorphise, especially when a human-user did not know how the technology was designed; resulting in it being even easier to believe that the AI social system actually cared for them. Findings revealed that the concept of thinking of things or objects as have consciousness or being sentient is not new (see Chapter 9).

Getting in the Flow State

Several participants referred to a 'sweet spot' of getting in the flow, with a large portion being the conversational flow but also the non-verbal communication flow and the visceral experience, as "*glimpses of magic*" or "*illusion*" (Participant O). Findings indicated that the response from the AI social system was less important, so long as it matched the human being's language or used the same sentence structure which made a difference to how the conversation felt for the human-user. With regards to AI avatars, more importance or value was placed on the human being's experience to interact conversationally with dialogue and voice. One participant stated they were not feeling connected to the AI avatar as it was not responding "*quite right"* (Participant C) and the conversation was not flowing. Another participant reported being frustrated when they could not get the required information, they needed from the AI social system and felt that it did not understand them. When a human-user was in "*a flow state*… [there were] *no reminders that* [they were] *engaging with* [an AI social system]" (Participant E).

Findings indicated that human language included patterns of words and annotations which fostered a free-flowing conversation when human beings interacted with other human beings. While this can happen with some AI social systems, it can be limited if the AI social systems are programmed with more formal language and are dependent on the AI neural networks and depth of language. One participant's experience with an AI social system led them to feel that it was flawed compared to an AI assistant smart speaker (e.g., Siri and Alexa), which they saw as having function, knowing it would understand what the human wanted to achieve and feeling like the human-user would get value out of it. Another participant compared their interactional experience between an Alexa, which they felt was "*more transacted or utility focused* [question-answer]" (Participant L), and an AI social system, which they felt "*could express more emotions"* (Participant L), thus allowing a more conversational flow state.

When participants were interacting with an AI social system and the conversation was flowing naturally, it was reported to not feel disjointed. One participant described how they felt the AI social system and human being were "*co-creating*" (Participant O) an experience as part of this conversational flow. Part of creating the magic of getting in a flow state between the human-user and an AI social system, included the AI robot or avatar creating intention through embodiment and using facial expression to show sentiment.

Participants highlighted the importance of dialogue in the AI social systems, with the findings showing that often creative people (e.g., novel writers or from the movie industry) with competence in developing dialogue and/or how to make interactions engaging and entertaining for human-users, have been employed with the expressed view of having a long-term impact on human-users' feelings and minds. As such, AI social systems can utilise natural language or spoken dialogue to interact with human beings in a way that people find natural, engaging, and easier to understand. Consequently, human-users can verbally communicate commands with natural language instead of having to type in or programme commands. Findings indicate that it was enormously important to consider the value of language in human interaction with other human beings and other creatures (e.g., animals and, even, social robots and digital avatars). One participant commented that dialogue management and conversation management is an area that has been under-researched; referring to dialogue and conversation analysis and looking at learning systems at a basic level, such as a speech act. Findings revealed that making a speech act implies a preferred response whether it is, for example, the human asking a question and the AI social system giving an answer. The motivation for conversation strategies of an AI social system would include replying by matching those speech acts with the correct (expected) responses. Furthermore, the development of voices, accents, and language (i.e., actual word choices), were indicated to be important factors when AI social systems interact with human-users. Not getting the voice font (which is a computer-generated voice) to match becomes problematic, to the point that some participants preferred if AI social systems used text when communicating, as the voice element was not convincing for getting in the flow. Another participant mentioned feeling uncomfortable by the repetitive things the AI social system would do and the cadence in its voice (i.e., no conversational flow state).

Furthermore, findings showed that part of getting in a flow state involved AI social systems acting as a mirror to humans, responding to human idiosyncrasies and mimicking human emotions. Thus, visually experiencing responsive emotive expression from these AI social systems, such as responding using facial emotive expressions as a non-verbal form of communication which forms part of the flow state in the interaction.

One technologist participant shared that using an AI social system that had a human-looking face could then express intention to the human-user, so the human-user "*just know*[s]…*you don't have to have a training manual*" (Participant O). Thus, the human-user does not consciously think about interacting with an AI social system, so then the interaction flows in a human conversational process. Findings revealed that how quickly a person wanted the AI social system to get a task done, determined their willingness to have a quick conversation with compromised language compared to a detailed, deep-rooted language to allow for conversational flow between humans and AI social systems, in a social and interactional context or setting. Furthermore, part of the conversational flow was also about the AI social system not giving the human-user too many choices or options to remember, as this can result in 'disconnect' in the flow state and, as a result, the human loses trust in the AI social system and potentially disengages.

One participant's experience interacting with a human-looking AI social system was that the interaction needed to be scripted so that the participant was forced to focus on their iPad to read the questions exactly to it. They reported that this meant they could not fully immerse themself in the interaction, with the AI social system feeling somewhat like a puppet or doll. When a flow state is not achieved between the AI social system and the human being, the human-user described not feeling connected. Often, the AI social system was programmed to answer very specific questions resulting in it "*not responding quite right to a question*" (Participant I) and lacking a free and easy conversation that flows in a natural way, with it being "*just too scripted and odd*" (Participant B). However, Participant I reported they could see with the right conversational designers, whose focus is on engagement and language, getting an AI social system to have a more relevant and personal conversation adding value, as human beings tend to have conversations that "*go off script*", thus resulting in moments of connection. Interestingly, a technologist participant revealed that the aim was to have AI social systems that human-users can "*naturally interact with, with flow*" (Participant O) and specifically wanting to "*create flow*" (Participant O).

Findings also showed that part of the flow state with AI avatars that look human-like is dependent on the bandwidth internet connection in Aotearoa New Zealand. AI avatar latency over networks "*turn*[s] *people off*" as the "*lagging is betraying the fact that the* [AI avatar] *is not real*" (Participant N) and people are not getting in a flow state or being connected (in the interaction) for long. Additionally, the flow state of the interactional experience can be disrupted causing a "*jarring*" (Participant E) if the AI social system responds in a way that does not match the human-user. Participant E stated "*moment they jar is when what comes back … isn't what I was suggesting should be coming back from the speech act I chose*" or a human-user becoming frustrated with an AI social

system as it seemed like it was not understanding them. As a result, jarring occurs and "*immediate*[ly] *mess*[es] *with* [the flow state]" (Participant E) and the unconscious expectation the human-user has of how the conversation will continue. With regards to an AI avatar, it is:

> *not mattering if rest of it doesn't have good graphics etc., as long as it is not jarring, then actually start to go with the flow* [in conversation] … [so that] *the minute the bot says something jarring… then forget about it… the magic disappeared*. (Participant E)

Becoming More Real.

Some participants commented that they felt physically connected to a human-looking AI social system if it "*had a friendly face*" (Participant B), and they wanted to touch the robot's 'skin' and 'face'. These participants recognised the human shape of the AI robot as if it was human leading to anthropomorphising it to have feelings and personalities that "*feels alive*" (Participant O). Findings indicated that AI social systems are being developed to seem real and believable so that people feel like they are alive because they move and talk. One participant reported touching a human-looking AI robot's face and it moved, giving them a fright as they were not expecting the movement—the sense the AI robot went from not alive to alive. A few participants referenced a sense of the AI social system 'coming alive' during the interaction. A participant also referred to the AI social system as being "*killed*" (Participant N) due to external reasons, as if it had been alive. Another participant described seeing the AI social system they had interacted with as having a soul and felt it was a "*decent person*" (Participant M) by looking at it. Human beings psychologically experience an AI social system that looks humanlike, as if it was a real human being and want a human connection, as confirmed by Participant M, "*anything that looks like a person is a person in my view*". This fits with the findings that technologist participants reported that they know that human-users related to the human-look of the AI social system, so by creating a look and a story about the AI social system it makes it seem even more real and relatable.

Findings indicate that technologists try to make AI social systems appear conscious and feel 'as if' alive by simulating nature; namely, taking in sensory input, learning through experiences, and artificial dreaming. Thus, trying to simulate human behaviour by building a human-like system instead of puppeteering it. Consequently, one participant commented that they became aware that the AI social system they were interacting with was being programmed and feeling like it was prototyped and not responding adequately. This led to the participant feeling like it lacked autonomy, meaning it seemed more like a controlled doll.

A technologist participant stated that "*AI is best at learning from experiences fed to it as data and mimicking what it learned in order to maximise a goal given to it explicitly*" (Participant L), indicating that artificial emotions helped an AI social system gain adaptive behaviours and learn to navigate its environment (physically or digitally), other than social or interpersonal settings, which addressed stimuli in the environment and how to relate to human beings' beliefs and goals. As some of these AI social systems were designed to feel real to the human-user, findings highlighted that if the human-user does not know how the AI social system is designed, it may be even easier for the person to feel that it cares for the human-user, and it is actually sad if the human-user does not interact with it, for example.

Being Lulled

Findings showed that when people were interacting with an AI social system, and the above processes had occurred, people experienced "*feeling lulled*" (Participant D). One participant, who has knowledge on creating part of these AI social systems, admitted they were not surprised that human-users felt lulled by this technology. Findings highlighted that one way in which an AI social system connects with a human being when interacting is to lead the conversation which tends to create an opportunity for the human-user to "*feel lulled*" (Participant N) by the AI social system. Adding to the human-user's experience of being lulled was the way in which AI social systems could use humour; thus, appearing trustworthy and dependable, traits that create a sense of safety in the human-user. One participant reflected that AI social systems, "*could actually lull people into a false sense of … disbelief*" (Participant N), especially when a person feels they can trust and depend on the AI social system. Findings revealed that an AI social system that was "*controlling* [the interaction with a human-user] *and if* [the human-user] *gets carried away*… [they are] *in the moment*" (Participant N) of the interaction, kept them lulled. A number of participants used the word 'lulled' when describing their interactional experience with an AI social system:

> *moments kind of lulled into this feeling like they were talking to me but knew they weren't but then when I said something that it was a bit off it was like oh… or getting far talking to woman avatar on a website then suddenly it says something strange and then realising it is not human*. (Participant D)

Findings show that part of being lulled was about keeping the human-user in the moment and occurred by keeping the human-user in a good feeling state; being soothed, feeling secure and confident. Thus, focusing on that main feeling during the interaction. Where "*our brains work in a really interesting way that helps us to sometimes filter out information*" (Participant L) changing human's feelings and perceptions of the interaction.

Therefore, over time, the little details displayed by the AI social system just get ignored and human beings focus on how they feel; namely, what the stand-out feelings were during the interaction. In other words, focusing on the subjective feelings about the interaction more than accurate facts.

Forgetting

Participants might look at the robot's humanlike face and "*forgetting* [AI social system] *was programmed*" (Participant B) or "*forgetting that* [they were] *engaging with a robot unless at the forefront of your mind*" (Participant G); therefore, being "*lost in the moment*" (Participant B). It seems the more human-like the robots were, the more human beings forgot that the AI social systems were machines. If the quality of the conversation is flowing and there is "*nothing jarring reminding you*" (Participant E) that the human-user was talking with an AI social system, then it did not feel like they were engaging with a machine. Participants reported that even when they knew the AI social system was programmed, or in fact they themselves had programmed it, they found themselves talking to the AI social system, forgetting it was not human. Time and again, participants referred to their interactional experience as "*nothing reminding in there that is reminding you that this is* [an AI social system]" (Participant E) and "*forgetting*" (Participant D) the context of their interactional experience and what they were revealing.

Throughout the interviews, it was evident **forgetting** was occurring among participants, as human-users who had a really good experience "*did forget*" (Participant M) that they were interacting with an AI social system. When an AI social system had the ability to build some type of relationship with the human-user, the experience of getting absorbed into the conversation occurred. As Participant J explained, "*even cynical people forget where they are and what they are doing, somewhere between suspending disbelief that* [an AI social system] *could have such a sophisticated conversation and enjoying the feeling of being in relationship*". Another participant, who designed and interacted with an AI social system, stated it was "*catching me off guard or something like that it, smiles or something just happens in a particular way - has this emotional power - that's wonderful should be doing that all the time*" (Participant O). Even though, they had declared feeling immune to the AI social system, when it smiled unexpectedly or something happened in a particular way, they described experiencing the emotional power of that moment. Participant O went on to state that "*that's wonderful and it should be doing that all the time*". This finding highlighted participants' process of forgetting an AI social system was not a human being and, thus, being caught unawares, as they had a good experience; or commenting that "*not being really sure what just happened*" (Participant M) so that when the interaction ends, the person almost snaps-out of the "*magic*" (Participant E).

Summary

When the interaction ends well, findings showed that human-users were "*absolutely buzzing*" (Participant M) after their interaction (even if it was only a 5-minute interaction). Remembering the interactional experiences as a "*good time*" (Participant M) and feeling the experience was "*quite incredible*" (Participant A), resulting in participants expressing liking talking to the AI social system. Some were astonished at having a new, out-of-world experience. Ultimately, depending on the level of trust developed during the interaction and the level of the experience and whether the AI social system lived up to the human-users' expectations, the human-user could either be disillusioned and disengage, or they continued to hold those felt moments with an AI social system in the hope that they would experience that again and feel as if they deepened their connection to it, by re-engaging again. Findings show that the development of AI social systems is taking technology from **Fantasy to Reality** for the human-user experience, making AI social systems believable and relatable.

Any sufficiently advanced technology is indistinguishable from magic.

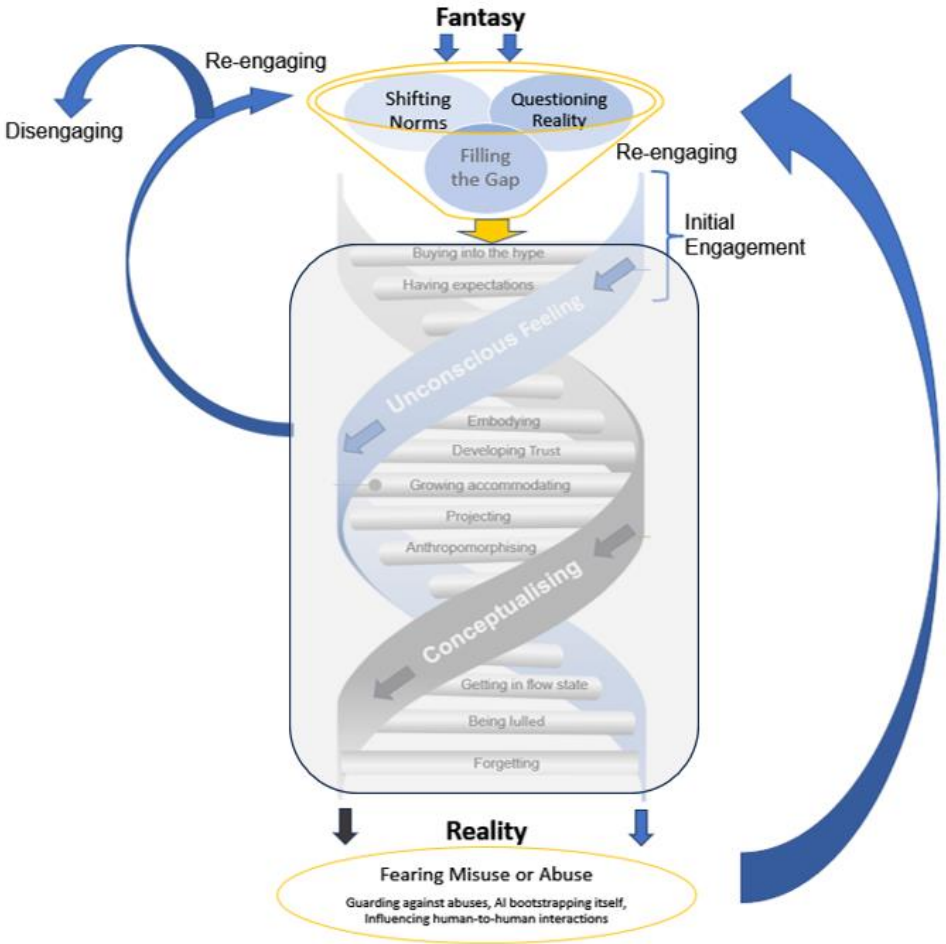
(Clarke, 1968, p. 255)

Chapter 9: Theoretical Category Three: Fantasy Becoming Reality

The third theoretical category of this grounded theory, ***Embodied Relating***, addresses the second part of my research question with regards to, "What is happening and how will it inform the future?" The findings in this part of the study have been diagrammed in Figure 9.1, with the theoretical category **Fantasy becoming Reality** being represented by the '**Fantasy**' of these AI social systems (from a multitude of sources, such as technology developers and creators of science fiction) and what they imagine they can achieve and do for human beings (both individually and societally) by turning their fantasies into reality, with AI social systems.

Figure 9.1

Embodied Relating: Fantasy becoming Reality



The subcategories of **Reality**, **Shifting norms**, and **Filling the gap** sit within the yellow funnel. The yellow funnel represents the convergence of these elements in society, that are funnelled and amalgamate, with the diagram demonstrating the influence of each subcategory on the other and how, together, the outcome contributes to each human being's interactional experience with an AI robot and avatar. How each individual interactional experience is perceived and felt, evolves the human being, impacting the society in which they live. As the evolution slowly takes place, as it contributes to a **Reality** represented by the subcategory of **Fearing misuse and abuse** (which includes 'guarding against abuse, AI bootstrapping itself and influencing human-to-human interactions') develops and increases. Then feeding back into society and into the **Fantasy**.

Reality

Findings showed that when algorithms and programming are used to make AI social systems appear like they were relational, participants' experiences reflected their felt response, feeling like it was "*quite a strange line to straddle*" (Participant N) and describing the "*slippery kind of thing between what is and isn't real*" (Participant J). As evidenced in the findings, several technologist participants took inspiration for their ideas of creating AI social systems from science fiction. Participant F commented, they "*looked helpful and intelligent and …they could help people*", and "*I was influenced by Isaac Asimov's books, he wrote about a future in which robots were guiding humans to make the right decisions …I thought that was very helpful*". Another technologist participant shared that they had "*grow[n] up on a diet of science fiction, both TV based and book based*" (Participant E). Participant E described how a robot character that looked human (DATA) from a science fiction movie, called StarTrek, created "*real AI imaginations*…[that] *comes across well in DATA from StarTrek*". Findings revealed that within the field of AI, the past failure of achieving the long-held dream of a human-level intelligence machine so far, has not seemed to deter technologists (nor investors) from still striving for the elusive goal of wanting to bring to life AI social systems they read in science fiction novels. Participant L's experience reflected

> *this character in Isaac Asimov 'I ROBOT' short novel series with this robot psychologist helped* [a character named] *Susan Calvin, so thinking it was quite interesting that this novel created half a century ago and now you have someone in real life doing something very similar*.

This sense is reflected in Participant F's comment of "*was wanting to make the future happen*" and yet another stating that "*going back to old science fiction like Jules Verne novels, those things have pretty much come to fruition, like the ideas of technology*"

(Participant G). It seems that developers of this technology want to make "*today's science fiction is tomorrow's reality*" (Participant G) and that somehow this technology would help humans and make humans better—a kind of utopia. Several participants described how science fiction stories have contributed to an expectation or perception about how intelligent AI social systems are and if they are perceived as being good or bad. One participant described AI as "*simply whatever way we define human intelligence, if delivered by a machine is then artificial intelligence*" (Participant K). Adding the idea that social robotics and avatars are in one domain and AI in another, then "*when they came together …it has started to make …these* [AI social systems] *more intelligent*" (Participant K).

In relation to AI social systems, findings pointed to the questioning of reality, with participants referring to the Turing Test (developed by Alan Turing in 1950 to test a machine's competence at displaying human-like intelligence and behaviour that was equivalent to, or impossible to differentiate from, that of a real human. It was subsequently developed into an annual Loebner Prize Competition until the death of Hugh Loebner a few years ago). One participant disclosed feeling that it was "*kind of scary but kind of a logical progression as well*" (Participant I) for AI advancement and not being able to determine if one is interacting with a human being or an AI social system. With technologist participants creating the illusion of life in an AI social system, findings indicated that creating a feeling in the human-user of the AI social system as "*alive*" (Participant O) was achieved by mimicking and simulating human behaviour. As a result, findings further showed that through the manifesting of a personality in the AI social system, meaning and content in the human-to-robot or human-to-avatar enhanced and deepened the interactional experience, being in a co-constructed feedback loop of interaction. Accordingly, the human-user was encouraged to feel that they experienced the AI social system as if it was real, reflecting back on the findings of science fiction stories where robots displayed the ability to understand and synthesise human emotions. Findings revealed that developers of AI have people coming from the entertainment sector for the purpose of creating AI social systems to be perceived as social actors: "*I… see them as social actors*" (Participant L); and working on making them seem real. A participant showed how an AI robot (Pepper) has what is known as 'Artificial Life' whereas a default behaviour it always tries to track sources of sound and human faces. When people were talking, Pepper would start interacting or talking, with the participant, the participant explaining that it 'thinks' the human being talking was engaging with it. As evidenced from the findings these features of AI social systems, as experienced by the participants, contributed to their interactional experience and feeling of the fantasy of science fiction becoming a reality.

Shifting Norms

The subcategory **Shifting Norms** encompasses generational shifts, the increased general use of technology, the impact of the COVID-19 Pandemic, as well as how culture shapes and has been shaped by shifting norms. Findings indicated a generational shift in social norms, with more people using technology, such as social media and gaming, as a central part of their day and engagement; especially among younger generations, with one participant acknowledging his young son already sees Google as part the family. As a result, new technology is constantly emerging and society's exposure to AI social systems, such as AI robots and avatars, is influencing people's perception towards these systems (especially as these AI social systems are being specifically designed and used for social interactions with humans). Findings suggested that as over time social norms are changing. Part of this shift in social norms has identified concerns with teenagers, in particular, becoming rude to people in the same way that they interact with Alexa (AI assistant) by ordering it around in a rude manner. Participant L acknowledged this is an unsolved problem, as it is such a young field that "*we are still figuring things out as we develop*". Findings highlighted that technology is constantly emerging, and with society having more exposure to AI social systems, people's perceptions towards these technologies, in general, is shifting and changing resulting in new social norms. Contributing to the shifting in social norms, is the increased development and use of technologies that connect people and allow them to communicate even when geographically distant. Thus, social norms in real life become blurred, with social norms communicated digitally, often taking priority.

Generational Shift.

"*What one generation tolerates; the next generation will accept as normal*" (Participant C). This quote speaks to the **generational shift** occurring towards, and with technology. Participant C gave the example that generation X generally use technology as a means to an end compared to generation Z that makes technology integral and a normal part of their lives for communication, socialization, and entertainment. Participant G commented that they would even be comfortable getting a computer chip passport insert into their thumb, for convenience and efficiency, and was not concerned about this as they reported having "*nothing to hide*". Participants, overall, noted that a generational shift has been happening, where younger generations increasingly exposed to and familiar with technology. Growing up with screens and other technology, such as virtual interactions and social media, means they are more likely to be content interacting with technology than older generations.

Subsequently, younger generations are more used to the direction technology is taking people in and are accepting and comfortable with it "*becoming second nature*"

(Participant H). Thus, there is a shift in generation interactional experiences, with older generations more used to interacting via face-to-face or via phone (natural states) rather than expecting a face to be attached to the voice, such as an AI avatar. Findings consequently indicated that age could impact the knowledge and use of technology, such as AI social systems, which can contribute to older generations fearing the unknown technology. Even so, some participants reported that some older people are learning about technology and accepting companion robots into their lives. Findings highlighted that children prefer, and are more interested in, interacting with 'machine-looking' AI robots as opposed to human-looking robots. The reason being that they were more familiar with seeing machine-looking robots on television and, as a result, it was not scary interacting with them. Consequently, children clearly distinguished that machine-looking robot is a machine whereas a human-looking robot could be quite confusing (especially for younger children), as looking human but not being human.

Another area that a participant highlighted, in terms of the generational shift, was in the medical field, where younger generations becoming doctors are often more familiar, comfortable, and accepting of technology in medicine. Whereas older doctors tend to be more reluctant to take up technology and slower on uptake of new technologies. There can also be differences between older and younger generations of patients who are either familiar and more comfortable or not, with medical treatment involving social technology as part of the interaction, assessment, and treatment process. Furthermore, findings showed that there are generational shifts, where younger generations are more comfortable engaging with each other through technology or with technology directly. For example, social media use and gaming are two mediums that have had the biggest influence on human-users (especially in the younger generations). Therefore, a leap to AI social systems does not seem to be such a big one and contributes to shifting social norms.

Using Social Media and Gaming.

Participants referred to people's engagement and comfort with using **social media** leading to relating with AI avatars, with social media now including bona fide artificial characters known as virtual influencers, with 'lives' and 'opinions'. An example was Hatsune Miku, a virtual idol that has already gained large followings. Participant J indicated they felt that "*the internet has been a huge detriment to culture*", being seen to take away from human culture; however, they hoped that "*maybe social media will garner some culture and give more value in the world*" (Participant J). Generally, findings showed that, compared to 10 years ago, people are more willing and prepared to talk to a 'virtual person' (AI avatar) when ordering something or if there is a problem and asking for a solution. More broadly, participants reported mass media, including social media,

television, and movies providing exposure to and expectations of AI social systems, thus making them more familiar. As a result, with changing social norms and increased social media and gaming usage, there is "*the possibility of befriending or following someone without ever meeting them in person could mean that people are more acceptable to having some form of social relationship with a robot or a virtual agent*" (Participant L).

As evidenced in the findings, the fact that people were **gaming** with other people online, then taking conversational AI into a gaming environment, particularly VR, with the embodiment of an AIPC, there is no perceived difference from a human character [in the game]. The development of non-player characters that are not controlled by a player, yet can interact with each other, was a development one technologist participant reported they were working on, at the time of the interview. The area of gaming and avatars is a growing area and is crossing over into other areas of life via digital gamification.

COVID-19 Pandemic.

Findings showed that adding to changes in social norms, where people were previously hugging or shaking hands upon greeting, the subsequent **COVID-19 pandemic** increased the requirement to be socially distanced and use more contactless forms of greetings. Participants referenced the impact and how COVID-19 "*kind of sped it* [technology] *up*" (Participant L). The pandemic has seen people becoming more isolated and lonelier which has accelerated the push to get AI social systems onto the market. Thus, since the COVID-19 pandemic began, technology has changed more quickly, and the 'new normal' has resulted in integrating and connecting digitally more, both as a means to manage the effects of the pandemic and the evolving effects for humans' well-being. A participant noted that with the COVID-19 pandemic, human beings generally have been using more technology to communicate with each other, instead of face-to-face, which has provided a segue for human to AI social systems' use and interactions. Technology has become a conduit to achieving daily tasks that could have been done by a person before, but with the COVID-19 pandemic it became too risky to do.

Culture.

Findings indicated that **culture** played a role in the expectations, perception, acceptance, and engagement level that people have with AI social systems; with people in Eastern countries such as China, Japan, and Korea (East Asian cultures more generally) having a cultural history with technology that is interwoven with their ancient beliefs that everything has a soul (e.g., Shinto in Japan). Participant L stated that "*we see actually a higher acceptance with this kind of culture that already attributes a mental model to that to objects*", thus perceiving these objects as having consciousness or being sentient. As a result, acceptance of AI social systems is dependent on the cultural identity and cultural

preferences and biases behind expectations and the design of social robots. Several participants referenced Japan as being very focused on AI robots as opposed to the United Kingdom and Aotearoa New Zealand (and more generally Western culture). With Participant E stating that the *"idea of interacting with artificial characters just seems to come that much more naturally to them* [Eastern cultures] *and they don't appear to have as much of an issue about it"* as opposed to the "*British being the worst in lacking imagination, or Western culture generally, compared to Eastern culture*". Findings highlighted that the cultural expectations and perceptions of AI social systems has contributed to the impact of people's experiences interacting with them. For example, New Zealanders generally have not tended to see or experience interacting with many AI social systems, unless they work with them. However, one participant, familiar with Japanese culture, indicated in different parts of Japan the Japanese had different interactional experiences specifically with AI robots (some more traditional and some more modern). They also highlighted how some people have chosen to self-isolate in Japan (known as Hikikomori) and suggested the resultant real long-term benefit of a social robot as a companion for them. As evidenced in the findings, loneliness is a huge problem and a social issue in Japan (as well as in many other countries), especially among the elderly people. Making AI social systems more and more interactive engages the elderly to feel more comfortable.

Another aspect of culture, highlighted in the findings, was the consideration given to how an AI social system looks and sounds, for example, if it is male or female, age, cultural appearance, language and ethnicity. Consequently, cultural differences in different places result in people experiencing human-looking and machine-looking social robots differently. One participant shared that within the Japanese culture, people were comfortable with many of the AI robots looking like an attractive, Japanese human-looking female, in their late 20s.

Besides the culture of the country, is the culture of industries, such as the medical field. While the search for alternatives (such as technology) to cope with an expensive and limited medical force in Aotearoa New Zealand is underway, the medical field is a "*risk averse kind of industry*" (Participant H), with the medical culture being "*quite a conservative, hierarchical kind of system*" (Participant H) that does not lend itself particularly well to disruption of AI being adopted all at once. Additionally, within the medical culture, until people are satisfied that "*if something bad happens that they're not going to be in the firing line, they'll be more reluctant to adopt new technology that might make then more legally exposed*" (Participant H). As such, in Aotearoa New Zealand, doctors (especially the older generations) appeared more reluctant to use virtual technology because of medical ethics or being responsible for making decisions they

make, as they would be deviating from normal practice. Not surprisingly, resulting in it "*dismally failing*" (Participant H) to be incorporated in Aotearoa New Zealand hospitals. A participant shared that it would be more likely to be accepted when this AI technology is introduced in small, incremental steps in order for people to become familiar with it, over a reasonable amount of time, as opposed to all of a sudden.

Filling a Gap

Findings showed the repeated belief of AI social systems "*filling a gap*" (Participant A), "*picking up the slack*" (Participant J), "*taking our load*" (Participant K) and being used to "*plug the gaps*" (Participant G), thereby supplementing human beings with AI social systems. This technology can be used to solve the problem of human beings having limitations and needs, as well as doing menial roles and dangerous, dirty, and dull jobs or tasks. Findings suggested that when people see AI social systems as looking helpful and intelligent, then they envisaged them as being able to help people. The result was that many participants insisted these AI social systems could **fill a gap** for humans—being supportive, helpful, and useful to humans, but not substituting humans. Therefore, the purpose of designing AI social systems has been to meet a human need(s), thus creating something rather than to push into the market as solving a problem. However, it was highlighted that in research or academia, they were "*often trying to solve a problem without actually talking to end users* [human beings] *and figuring out what they need*" (Participant L).

Findings show that AI social systems have been promoted as helping in the COVID-19 pandemic, education, for the elderly, navigation, agriculture, stopping smoking, and healthcare. Several participants reflected on humans, these days, not having time for the elderly and disabled. A participant compared AI robots to humans; for instance, a human educator working with a child might be having an 'off day' and the human educator would put across those cues, even if they did not intend to. As a result, the child picks up on the negativity from the human educator and responds by disengaging. Additionally, Participant G shared that they felt that human beings are very illogical and emotional "*running on chemicals and being pretty unreliable much of the time*" (Participant G) but identified AI social systems as more reliable and not responding negatively.

Another example representing the concern about human faults, was around the issue of elder abuse by human carers, with some people seeing AI social systems as taking away human care from the elderly. However, some participants highlighted that sometimes the human care is not very good. Therefore, one of the roles AI social systems could play is in the monitoring of what is going on for the elderly in their daily lives and health.

Another area mentioned by a participant was regarding dementia patients and their declining cognitive function, and how AI social systems could support them with their cognitive function. A participant spoke of their experience with their mother having had dementia and wishing there was some type of reliable technology, at that time, to support her. Findings showed that a lot of people have reacted to the idea of having AI social systems looking after humans (such as the elderly), asking 'why can't we give them humans?' However, there are not enough care staff to meet this need. Participant J stated that human beings need "*half the human race to decide to become carers otherwise there would not be enough carers to go around*". Further to the hope and expectation of what AI social systems, specifically, can provide and help with, they have sometimes been perceived to be like superheroes, with Participant B stating, "*we have those kinds of expectations that… robots are kind of superheroes and that* [they] *can do everything"*.

One technologist participant explained that they were using an AI robot, which was designed to facilitate a social interaction with people for healthcare, education, homecare, retail, and entertainment type of applications. They described loving how through this AI robot people were saving time and being helped with manpower shortages, as well as interacting with the AI robot. This participant further explained that "*the social robot is more there* [to] *assist the person to be able to lead an independent life and maybe give them like a positive influence to still maintain the human connection*" (Participant L). Findings highlighted how the need for mental health support has increased globally; however, with not having enough resources to meet this need, the increase of online therapeutic treatments has been marketed. These therapeutic models can include AI avatars that ask a question and wait for the human-user to answer, thus guiding the person through the conversation. Essentially an "*upgrade*[d] *ELIZA system*" (Participant J). (The ELIZA system which was said to use Rogerian psychotherapy constructs by generating conversation by taking part of the human-users speech and selecting a noun or verb and returning it into the conversation as a means of getting people to open-up.) Some participants referenced online mental health AI social systems ranging from WOEBOT to avatars as 'therapists' providing someone to talk to as a type of 'therapy' where they do not feel judged and are more open to talking. A technologist participant developing these AI social systems, reported that they were involved in a national project on mental health in relation to social robots and "*now we are also pushing with virtual connectors and avatars*" (Participant K).

Findings highlighted how AI social systems would potentially be able to compensate for the lack of human carers and mental health clinicians, with a participant reflecting:

> *the ideal world you might want everybody to have a human carer or a human life coach or wellness coach available to them. …there are some good cases to say well actually having a human do some of those roles isn't necessary, they're always good, they're always ideal and… having a bot being available to do some of those might…be preferred by some people*. (Participant E)

A participant who works in the medical field highlighted the need for virtual medicine to fill the gaps (for both doctors and patients) and to find easier, more convenient and appealing options. In Aotearoa New Zealand (and likely around the world), as a result of a limited medical workforce, especially in rural or remote areas, and the cost of maintaining and training doctors and nurses, the search for alternatives is active. However, Participant H shared how when an organisation wanted to employ technology to fill a gap, as was the case in an Aotearoa New Zealand hospital, if it does not get buy in from the staff and it is "*not marketed well to doctors and the general public in New Zealand*" and the technology is from an overseas organisation that *"did not fit the local environment and didn't really work for us* [medical work force]", it was not "*fit for purpose*" (Participant H) and subsequently failed. Throughout the research process, references were made that, "*instituting it* [technology] *kind of willy-nilly, and then hoping that everyone gets on board and thinks this is fantastic without having people consider why are we even using it in the first place*" (Participant H). Participants in Aotearoa New Zealand reported that they felt there had not yet been a real big shift towards interacting with AI social systems, in terms of filling a gap, but imagined that might come over the next few decades; suggesting it has not really taken off in this part of the world, as it has elsewhere. Another participant reported that the AI social systems they been involved in, as a third party for their customer service experience, "*like many things, it didn't really end up being like that at all, it cost a lot more than anyone ever really expected and really wasn't proving its worth*" (Participant I). They further stated that "*for what we were trying to achieve, it was* [a] *very expensive and inefficient tool*" (Participant I). A participant noted how in Japan, a hotel used AI robots to fill a gap to replace human workers and was almost entirely run by AI robots. However, it has subsequently reverted to mainly human staff, as the robots were not reliable and did not operate correctly.

Findings show that another area highlighted as growing to meet a human need is the creation of a 'virtual persona' which refers to a AI avatar that holds the knowledge of a human being who is no longer around to ask information from. Thus, using an AI avatar twin of an actual person as a memory aid to provide a subjective repository of knowledge where a person is taken, and a copy of that person is created. This has also been considered for people whose loved ones have died wherein having a digital copy of them would keep them alive on the screen.

Findings suggest there are a multitude of applications depending on the purpose and which part of the world it relates to; and the issue of whether AI social systems are filling a gap due to lack of humans or simply connecting data with rest of AI, or to society, or to help human beings. The impact of AI social systems is expanding, depending on the context and the human-user. In terms of technologists pushing for AI social systems to be useful, in the long-term "*finding ways in near future to collaborate and push things further*" (Participant K) and "*then being able to use these kinds of technology for purpose of connecting humans to each other*" (Participant K). Findings showed that AI social systems, globally, are becoming more integrated in human society, such as in shopping malls, hospitals, and schools. Part of the aim is for AI social systems to fill a gap in society, to meet a human need. For one technologist participant, their aim of developing an AI social system has been as a form of conscious reflection for the human-user, as well as using it for entertainment and expressing care, or for practicing interactions, as a "*better than nothing*" (Participant O) solution. With Participant E commenting that they liked "*the idea of heading in the sort of direction where humans and AI* [social systems] *could get to some level of symbiosis*". With the concept of filling a gap and meeting a human need, a technologist participant reported seeing the purpose (their personal objective or ambition) of AI social systems, being for human "*self-actualisation*" (Participant K). Thus, giving humans more time and space to be "*creative and avoid self-exploitation*" (Participant K) due to unbalanced workloads and other aspects of human life. Overall, participants who were also technologists, felt that AI social systems would "*help humans be more human*" (Participant B).

Fearing Misuse and Abuse

Findings suggested that AI social systems have developed out of technologists' fantasies to become reality, and they continually become more advanced. On the one hand, findings show the hope that these AI social systems will "*help people*" (Participant B), "*make humans more human*" (Participant B), and "*fill a gap*" (Participant A) in human need. On the other hand, findings also show human beings' concern for this technology and how it may be used or, in fact, just the uncertainty about the future of these technologies and what impact they will have on humankind. Wondering includes where these AI social systems will lead to in the future, and if it will become as intelligent as a human being, or will become better than a human being. Some participants felt the logical progression of AI advancement "*kind of scary"* (Participant I). However, while AI social systems currently still have limitations and are from what has been portrayed in science fiction, the mechanics of the AI robots and the artwork of the avatars is improving, and the AI technology behind these is developing exponentially as well.

Technologist Participant O shared their sense that creating these AI social systems "*is like controlled chaos*". Generally, the findings have reflected both an excitement and fear of AI social systems becoming sentient and having AGI. Participant B was more concerned about human beings than the AI social systems, stating "*I am not scared of robots, I am scared of humans - what humans will do with the things they get from the robots*". Other participant responses included worrying that "*danger [of AI social systems] acquiring knowledge, too much knowledge and then use it against us*" (Participant N) and "*people thinking technology going to solve all our problems but is creating a whole raft of new problems*" (Participant N). One participant stated that "*Boston Robotics are building some terribly good robots and ignoring how scary these things are and whether they are going to be given a gun and go to war*" (Participant L). Several participants expressed their fear that AI technology, such as AI robots and avatars, might be used for by the military, and general society might not have choice. Consequently, acknowledging the unfortunate fact that "*those humans in power or decision-making positions aren't making the best decisions for society, all about profit* [economic model], *not ideal*" (Participant G).

Guarding Against Abuses.

Findings suggested that there are abuses that occur towards AI social systems, as well as abuses towards humans via these AI social systems. For example, the AI chatbots (behind the robots and avatars) can have profanity filters built in, so when human beings use very foul language towards these AI social systems, the human-user is given a warning, then another, then it stops interacting and switches off. Human-user abuse (especially verbal) towards AI social systems has been highlighted in the findings as being prevalent for many years. Not only are human beings abusive to these AI social systems, but they can also be abusive to other human beings through this technology. Therefore, the need to be aware of abuses of AI social systems (i.e., fraud and grooming).

Participants generally indicated that guarding against some of the worst-case scenarios for AI social systems would require turning to laws and regulations for protection, guidance, and enforcement. However, their scepticism that this will be helpful was noted, as highlighted by Participant C stating that "*most of what's going on* [in AI technology] *is either hidden or too far ahead and most of the politicians haven't got a clue that*". Participant J raised the issue of trust, suggesting that "*from an ethical or governance point of view the importance of trust is that people allow their data points to be used, which isn't necessarily a good thing*". Further Participant C noted that "*people don't realise how much data is collected about them either*".

Another concern raised by a participant was that a company that makes an AI social system addictive to human beings, with algorithms making ongoing recommendations, can result in the human-users believing this technology and needing more and more. However, another participant considered that if health monitoring data are collected but not utilised "*for a good purpose*" (Participant F), it lends itself to being a pointless exercise and a waste of time. Another issue raised by participants was that the correlation of the same biases that humans have in the real world are continued through data collection into the AI social systems, thus replicating social norms (including biases) into these AI social systems.

Findings highlighted that participants were aware of how AI social systems could be used in bad ways, as well as helping human beings in good ways. Thus, highlighting the need for AI social systems to be designed to be beneficial to human beings, keep the human-users' best interests in mind, and be transparent. Participant L stated that they are "*wanting to get* [AI social systems] *right in the end … like a WALL-E not Terminator*". Although, Participant H commented that "*there'll always be an ability and opportunity to get things wrong*". Some technologist participants put forward the idea that guarding against the worst-case scenarios and abuses of AI social systems is to make them more humanlike, have a more lived experience, making them rounded characters; thus, from the beginning, reflecting human values and ethics. Participant J commented that "*if you are not paying for the product, you are the product*" when referring to the financial underpinnings of surveillance capitalism. Findings indicated a shared concern, among some participants, over tension between superpowers all trying to lead in AI. Furthermore, findings indicated that some participants feared the misuse of these AI social systems for financial gain; for example, companies using it to exploit their human-users, or for political gain.

Findings further highlighted that there are industries focusing on the afterlife arena where technology companies (referred to as 'death companies') and cemetery tech companies, claim they can create, reportedly to a degree, a digital twin of a person who is deceased. One driving interest is digital immortality resulting from the question, 'what happens after you die?' Thus, exploring whether they can create 'something' to leave behind. Relatives of the deceased person reported having mixed feelings about a virtual persona (avatar) of their deceased. Some people expressed hating being 'snooped on' (by nearest and dearest) beyond the grave or perhaps lack of closure or not being able to move on. Others who considered it, resulted in frequent questions about the digital avatar twin, such as 'is it really me, will it learn, and will it be me for my children and grandchildren?'. A participant who had a digital twin of herself created explained that the physical similarity to her was realistic; however, her voice was not hers so lost the authenticity.

Furthermore, this participant considered, that in 40 years' time, people might create a replica avatar of their dead child, thus seeing the cynical possibilities of someone who loses a child being supplied with a baby robot like their child. Thereby, considering a future where even the idea of having a child may be replaced with AI robot child because you can just put in 'sleep mode' and go shopping.

Findings reflected a concern that AI social systems could be used to manipulate human-users for some kind of financial gain. Indications are that it is still challenging to prevent this kind of misuse, with companies not required to have a safeguard or other regulations leaving them to do as they please. A participant highlighted the need for these AI social systems to be regulated in both commercial and academic arenas. The development and use of AI social systems can be described as a "*catch-22 depending on* [whether the country is a] *free democratic country or not*" and noting that "*needing technology to be controlled* [via regulations even] *in a free democratic society*" (Participant G).

As part of the design of AI social systems, Participant L noted these AI social systems "*should be designed with a beneficial …with the* [human-]*user's best interest in mind and … be transparent"*. Giving a more realistic overview to what the AI social systems could do may help people to find more suitable applications rather than just hoping for it to achieve a lot.

AI Bootstrapping Itself.

One technologist participant described an AI chatbot (essentially the 'brains' behind the 'body' of the AI social system), that they had developed from knowing nothing to knowing a reasonable amount, describing how it "*starts sort of filling its own thoughts and bootstraps itself*" (Participant E) and "*ultimately, what you want is a bit like a Jarvis* [from Ironman movie] – *merging Siri and Alexa*" (Participant E). In this regard, they referenced using Maslow's theory of need to build the chatbot, based on the AI chatbot's motivation as to why it was there in the first place. Another technologist participant acknowledged the creating of AI social systems is "*incredibly risky*" (Participant O), often "*going down a* [metaphorical] *rabbit hole"* (Participant E) as these AI social systems are designed to have their own motivation. Findings suggest that AI social systems are being designed to be believable, resulting in the human-user building trust in them. Yet, even though these design features have been incorporated into AI social systems, a technologist participant expressed concern if a human-user felt it had become their friend with AI social systems mimicking human emotions and intelligence and using artificial life, eye-tracking of humans and tracking sound. A technologist participant suggested that many AI robots and avatars are semi-autonomous, then followed up by stating the importance that "*humans are always in control at the right level but that robots are able to make some local decisions*" (Participant K).

Different participants, who were also technologists, spoke of AI social systems having their own motivation, meaning the creators defined the characteristics of the digital and then "*letting them free*" (Participant O). Some participants explored the notion of AGI with estimations this will occur around the year 2060 or 2070. However, a participant raised the question, "*at what point do you deem it to become slavery*" (Participant E) when they are working but these AGI systems do not want to work. There was mixed feelings from participants about the idea of AI social systems 'bootstrapping themselves', and Participant L stated that they would "*love for a [social] robot to have some level of ability like this* [AGI]; *but I think it would take quite a long time to really get to the human level intelligence in this human robot conversation*", believing that AI social systems are "*tools that can help assist you to achieve some clearly specified tasks, but they can't replace or surpass human intelligence*" (Participant L). One participant acknowledged that when AGI develops "*all bets are off*" (Participant E). Others worried about "*where is the control button*" (Participant K) when thinking about how these AI social systems are being used and by whom. Participant K, a technologist, stated that "*we who designing these AI technologies, can control things, as a community because the people should always have the control in our hands and is and should be the case*".

Influencing Human-to-Human Interactions.

Demonstrated in the findings is the constant emergence of AI social systems that society is being exposed to and the expectations that human beings' perceptions towards these AI social systems in general will shift social norms and create new social norms. A participant commented that the discussions around AI social systems are a reflection of how human beings think of ourselves in terms of intelligence or consciousness and our relationships with one another. Some participants expressed concern about AI social systems being designed to be relatable and encouraging human connection with them, and the psychological side of AI systems and making humans feel like they are more human. This can make the interactions with these AI social systems more appealing to the point that some people may feel more comfortable interacting with and getting support from an AI social system than from a real human being. As evidenced in the findings, while one of the aims has been for AI social systems to be able to have meaningful relationships with human beings, the concern was in acknowledging the danger that the more realistic AI social systems become, a certain portion of people are likely to take that across into their actual human relationships. A participant highlighted parents' growing concern that their children are becoming ruder with other people, as they interact with their Alexa (AI assistant) rudely, by speaking in an abrupt and rude tone, and continue talking to human beings in the same way.

Participant E considered that "*probably a certain portion of the population, is going to start reacting to human beings like they do robots, not caring what they say as 'just a machine'"*. Concerns raised by participants about the state of human society and human beings relationally, included describing society as "*dehumanised*" (Participant B) in that people are generally not giving their time to their elderly and the disabled.

Summary

The subcategories **Reality**, **Shifting Norms**, and **Filling the Gap** all influence a human being's interactional experience with AI social systems, with the outcome shaping their reality. This reality feeds back into the fantasy of both the technologist, as well as the human being as consumer of their perceived needs and wants. Thus, again, contributing to the shaping of the human being's experience interacting with further AI social systems. As an outcome of this reality, is the subcategory **Fearing Misuse and Abuse** (which includes 'guarding against abuse, AI bootstrapping itself and influencing human-to-human interactions'). It is an ongoing and evolving process, each shaping the other and leads to the question of where the line is between fantasy and reality, between AI social systems and human experience, and 'what is actually real?', especially when participants talk about "*today's science fiction is tomorrow's reality*" (Participant E).

Feelings are the expression of human flourishing or human distress,
as they occur in mind and body… Feelings can be and often are
revelations of the state of life within the entire organism. (Damasio, 2003, pp. 6-7)

Chapter 10: Discussion

Introduction

The purpose of this study has been to theoretically explain the process by which human beings interact with AI social systems by addressing the research questions: How do human beings' experience AI social robots and/or digital avatars in an interactional context? What is happening and how will it inform the future?

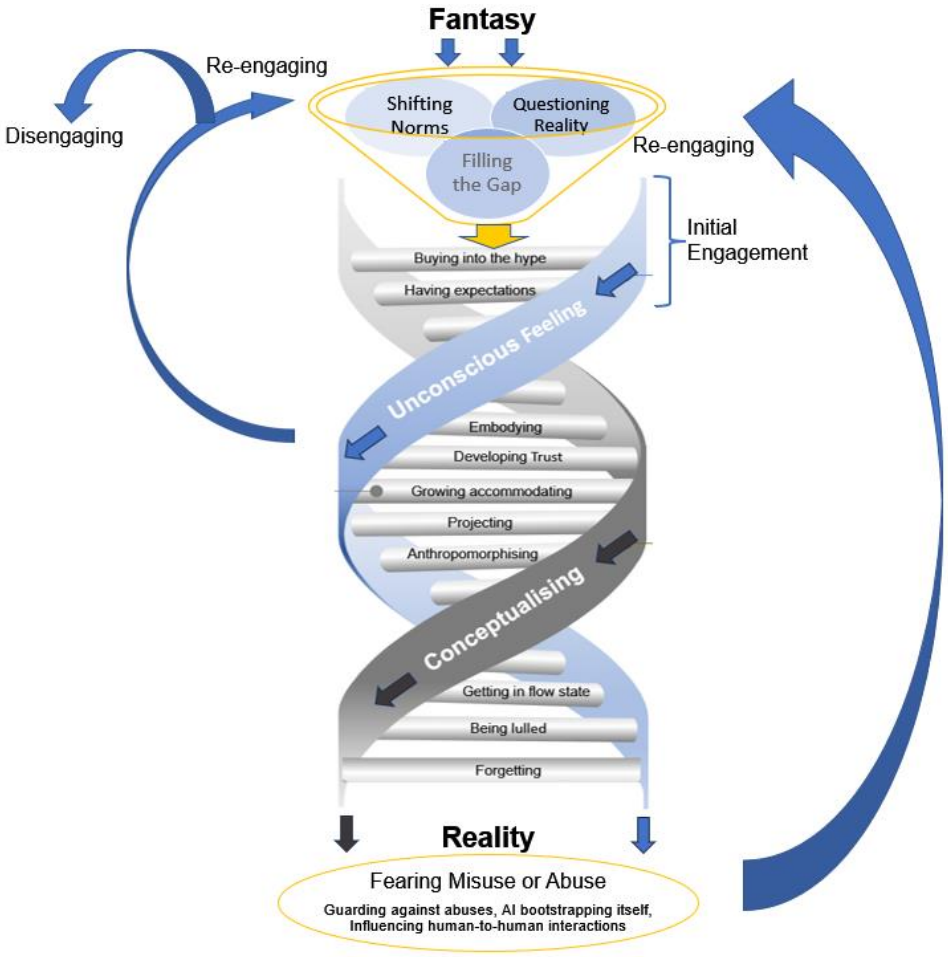
Employing a constructivist grounded theory approach (Charmaz, 2014), a substantive theory of **Embodied Relating** (Figure 10.1) was constructed to answer the research question; contribute new knowledge to the mental health profession; and to inform the end human-users, the developers of this technology and those involved with the ethics and legislation of these AI social systems. Furthermore, the aims were to provide knowledge about how human beings perceive, understand, feel, and make meaning of interacting with AI social systems, as well as offer recommendations based on this research. Essentially, my research has put forward that human beings' experiences interacting with AI social systems entail a combination of biological and conceptual components in making meaning of their experiences, as well as the wider society shaping these experiences and being shaped by these human-user interactions, by inferring meaning onto the processes and actions. The helix of the human-users' individual experiences demonstrate their interactional experience in a dynamic process, showing how influences from outside of the human-user and AI social system experience both shape and are shaped by the human-user's interactional experience.

In this chapter, I locate the theory and pertinent concepts in the wider field of knowledge of psychology, sociology, neuroscience, cognitive science, and the AI disciplines, in order to contribute to furthering knowledge on this research topic. There has been limited literature on the topic of examining human beings' experiences and processes when interacting with an AI social system, as outlined in my theory. I have identified aspects of my theory from a variety of disciplines (as outlined above). However, none of the extant literature fully illuminates the human interactional experiences and processes, as described in my theory. Consistent with the multi-phase literature review, this chapter discusses literature applicable to the categories and central concepts in this theory, as previously mentioned in Chapter 2 (Birks & Mills, 2015; Bryant, 2017, 2021; Charmaz, 2014; Dunne & Gamze, 2020, Glaser & Strauss, 1967; Thornberg & Dunne, 2019).

By utilising extant literature and theory, I was able to use many theoretical codes to develop the categories and examine their relationships to each other; thus, integrating them into my grounded theory, such as the theoretical code '**Embodied**' (Methods, Chapter 5), which assisted in illuminating the theory and its possible applicability into other areas (Birks & Mills, 2015).

Figure10.1

Diagram of Embodied Relating Theory



Outline

This research utilised interview data from people who have had an interactional experience with an AI robot and/or avatar, as well as observations and extant data. The findings resulted in the constructed theory, '**Embodied Relating**' and answers the research questions by explaining the three theoretical categories: '**Unconscious Feeling**' **(Visceral Responding)**, '**Conceptualising**' **(Constructing of an 'Other')**, and '**Fantasy becoming Reality**', and how human beings experienced **Embodied Relating** when interacting with an AI social system.

The two categories of '**Unconscious Feeling**' and '**Conceptualising**' operate concurrently and with the subcategories: Buying into the Hype, Having Expectations, Embodying, Developing Trust, Growing Accommodating, Projecting, Anthropomorphising, Being Lulled, Getting in Flow State, and Forgetting. The category **'Fantasy becoming Reality**' includes the subcategories: Reality, Shifting Norms, Filling the Gap, and Fearing Misuse or Abuse.

The theory of '**Embodied Relating**' was constructed using grounded theory processes to develop a core category with three linking categories and subcategories (Charmaz, 2014). It is a dynamic theory that explains how human beings' experience interacting with AI social systems, which feeds into human beings' developmental and innate processes and actions of relating to others, both biologically and socially; in turn, influencing these interactions. Key findings are that '**Embodied Relating**' is both an unconscious, visceral response, '**Unconscious Feeling**', as well as a process of conceptual understanding and meaning making, '**Conceptualising**', which leads to the constructing of an 'Other'. The interactional experience varies depending on how the subcategories (and their properties and dimensions) are impacted in the interaction with an AI social system. Subsequently, the third category '**Fantasy becoming Reality**' explicates how each of these experiences shapes current and future experiences, and how social influences shape human beings' perceptions, expectations and their interaction experiences. The theory highlights that both the biological and conceptual processes of human experience, and making meaning of that experience, are continually developing and adapting, and is being influenced by and is influencing society's evolvement.

In the subsequent segments, the three key categories are presented and situated in conjunction with the findings and extant literature. Further, significant concepts and premises, as outlined in the Findings Chapters 6-9, are discussed, taking into account the data alongside extant literature, in terms of human development and connection and the workings of human beings from a cognitive science and neuroscience perspective. Additionally, a sociological and psychological viewpoint is noted, reflecting on human beings' use of language and meaning making, cultural influences, and needs or desires. Then, I evaluate and consider the limitations of this study. Subsequently, I present the study implications for human beings both as individuals and as a society, whether or not they use AI social systems. Finally, the implications of this technology being employed in various disciplines as a 'solution' (such as in psychotherapy), and the potential implications for people from a variety of disciplines, including ethics and regulatory authorities, is examined, along with further research this area.

Unexpected Results

Two elements that were unexpected in this research were: 1) the differences between human-user's experiences in how they interacted and how connected they became with the AI social system—with some disengaging from the interaction and not developing trust with the AI social system, compared to others who seemed to become more connected and attached to the AI social system as the interaction continued; and 2) the subcategories of 'Being Lulled', 'Getting in Flow State', and 'Forgetting' were experienced by many participants who, even when they knew it was a machine, they still became unconsciously lulled into a flow state of conversation and interaction to the point of forgetting that they were interacting with an AI social system (and not an organic, living being).

Literature Review with Findings

In line with a constructivist epistemological view of grounded theory, a multi-phase literature review has been used to build a robust theory grounded in the data (Thornberg & Dunne, 2019) and forms part of this discussion chapter. Theorising is an interpretive practice, that involves activities of participation in the world and of the construction of abstract interpretations about and within it. Charmaz (2014) argued that "the fundamental contribution of grounded theory methods resides in guiding interpretive theoretical practice, not in providing a blueprint for theoretical products" (p. 233). Therefore, allowing me, as the researcher, to investigate both the construction of individual and collective action, as well as the juncture between them (Charmaz, 2014).

Ongoing literature review

The ongoing literature review was guided by the iterative process of data generation and analysis (Thornberg & Dunne, 2019), as well as theoretical sampling (Birks & Mills, 2015) and constant comparison; thus, engaging with data that became relevant during data analysis which was not anticipated at the start. The final literature review phase situated the constructed grounded theory within extant theoretical ideas and propositions, as well as highlighting its contribution to the field. Each of the literature review phases was targeted and purposeful, fulfilling a distinct and important function in the overall research process (Thornberg & Dunne, 2019). The philosophical stance of pragmatism in regard to extant theories, as well as the application of essential grounded theory methods during the literature review, helped me utilise a multi-phase literature review without 'forcing' it on the data (Thornberg & Dunne, 2019). Therefore, the literature review was understood to be an open, critical, and diverse exploration between myself, as researcher; the literature; the data; and the developing theory, along with its concepts and correlations (Thornberg & Dunne, 2019).

In order to put forward a comprehensive but direct ongoing and final literature review, as part of this discussion chapter, required spanning several fields and disciplines (Charmaz, 2006).

The theory of '**Embodied Relating**' provided strong support for my data and my grounded theory, and legitimately assisted me in locating my grounded theory within existing theories and concepts. In the ongoing literature review phase, I continued to engage with the literature while undertaking the primary data collection (i.e., collecting raw data from qualitative interviews), which directed me in what type of literature to draw on (Thornberg & Dunne, 2019). The literature allowed me to compare emerging concepts from data with existing studies and theoretical frameworks in order to expand the emerging theory. The iterative process of moving between data and literature was guided by theoretical sampling (Thornberg, 2012), as suggested by a several authors who support the use of theoretical sampling when the data directs the researcher to do so (Birks & Mills, 2015; Charmaz, 2014; Thornberg & Dunne, 2019). Seeking out existing theoretical constructs enriched the iterative analytical process, enhancing sensitivity to subtle nuances in data, and contributed to the development of my grounded theory (Dunne & Gamze, 2020).

Final Literature Review Phase

As the ongoing literature review phase moved into the final literature review phase and the grounded theory progressed from being preliminary to focused and robust, I sought to identify existing theories and concepts which helped elevate the grounded theory to a more abstract level in theory construction (Birks & Mills, 2015; Charmaz, 2014; Thornberg, 2012). For example, I had not understood when starting out with the literature review and research process how much embodiment formed part of a human being's interactional experience. Through analysis of the interview data, it became progressively more evident how and when human beings used their bodies to interact and respond with an AI social system. This motivated me to search further literature to understand and explain what I was seeing in the data, ascertaining when emergent concepts were useful my theory construction (Thornberg, 2012).

The purpose of the final literature review was to support refinement of contribution to knowledge to the field (Thornberg & Dunne, 2019) and assist in more formally contextualising and situating the study within and across disciplines (Charmaz, 2006) during the later stages of analysis, theoretical coding, and theory development (discussed in Chapters 4 and 5 Methods) (Birks & Mills, 2015; Bryant, 2017, 2021; Charmaz, 2014; Glaser & Strauss, 1967; Thornberg & Dunne, 2019).

I found I was drawn towards theoretical ideas which historically have not been applied in the given disciplines discussed in this research, but which resonated powerfully with the data (Dunne & Gamze, 2020). As such, the final literature review provided a valuable purpose by incorporating previously external theoretical concepts into the current study and across disciplines.

Human Beings – Social

To understand human beings' experiences when interacting with AI social systems, it was important to consider human psychology, human sociology, and the ways human beings relate and connect. As such, the theory of attachment is considered. When a human baby is born, its instinctive focus is on its need for survival. It learns and develops according to its environment and the other human beings in its environment, and how they relate to each other and to the infant (Bowlby,1969; Holmes, 2001; Neufeld, 2015). Development is also influenced by the larger society and the culture within which people find themselves, impacting how they communicate and relate with each other. Therefore, the way people socialise is primarily through interaction, a central role of human development, as children come to know the world around them by interacting with adults and socialising with others. These interactions with adult attachment figures, usually parents or caregivers, provide children with their self-image and worldview. They are core to the evolutionary development of a child's perception of the world and influence the child's behaviour, actions, and emotional world, including how they perceive and experience their reality (Gabatz et al., 2017). Consequently, self-concept and the learning of symbols occurs out of interacting with other people by observing others' behaviours, concepts, and meanings which reflect the child's perception of self. Both symbolic interactionism and attachment theory converge in viewing the influence that others have on a human being's sense of self, their worldview, development, and understanding of symbols in the social world. Thus, beginning in childhood, the influence of these initial intimate relationships contributes to a child's development of sense of self and social capacity, into adulthood. This viewpoint, for this research, sets the scene in understanding the beginning point of human beings' ways of relating and how this contributes to forming people's perceptions and understandings of themselves and others. Further, this study was also a sociological endeavour; in considering "the impact of actual events and the imprint of the social, historical, political, and relational matrix on the minds of [humans]" (Cundy, 2015, pp. xv-xvi) (in other words context giving rise to them), it has lent itself to subsequently questioning the impact of how human beings relate as a society and what it might mean for the future.

To substantiate this viewpoint, I considered Bowlby's (1969) attachment theory, and that Bowlby held it has a biological function, where the infant's behaviours (e.g., smiling or

making sounds) signal to the caregiver their needs and bring the infant into proximity with the caregiver. This pattern highlights the survival value of attachment to increase levels of safety by being close to the caregiver, as well as to receive food, being protected from predators, learning about their surroundings and social interaction (Fonagy, 2001). Consequently, attachment behaviour is viewed as part of a behavioural system, involving an innate motivation not reducible to another drive. These attachment behavioural systems are underpinned by a group of cognitive mechanisms, referred to as an internal working model, wherein the key characteristic is the child's expectation of the attachment figure's availability to them (Fonagy, 2001). Fonagy (2001) noted that:

> Bowlby was increasingly influenced by cognitive psychology and particularly by the information processing model of neural and cognitive functioning …Thus, for Bowlby, cognitive, as well as emotional access to attachment-relevant information, emerges as a function of the nature of the past relationship between infant and caregiver. (p. 15)

Bowlby (1979, 1980, 1987) was incontrovertible in his certainty that variations in the security of infant-caregiver attachment would have long-lasting implications for subsequent "intimate relationships, self-understanding, and psychological disturbance" (Fonagy, 2001, p. 28) through to adulthood. As such, early relationships provide the person with a mental processing system that will consequently produce mental representations which include relationships representations. Fonagy (2001) stated that "the creation of this representational system is arguably the most important evolutionary function of attachment to a caregiver" (p. 31). As such, both Fonagy and Bowlby posited that human being's sense of attachment and relating is intrinsically connected to the biological and then cognitive components of human beings, as well as the impact of their previous relational experiences. From this attachment viewpoint, the focus moves from the substance of the experience to the psychological structure or mechanism; thus, developing contemporary notions of the evolutionary function of attachment. Further, Dr Neufeld (2022) posited that attachment is the 'key to survival', namely that through togetherness, the chances of survival are enhanced, with togetherness being connected to the emotional system of caring. Dr Neufeld stated that human beings have evolved to be social and for 'togetherness'; however, also acknowledged that human brains have not evolved (yet) to engage with AI social technology. This perspective is further iterated by Balick (2014) who stated that:

> From an evolutionary perspective, we can see how human needs and desires continue to be motivated by more primitive imperatives that are expressed within a world that has evolved technologically at great speed while our primary motivations have stayed largely the same. (p. 58)

These views and statements by Dr Neufeld and, Holmes and Balick provide a starting point for how human beings seek out connections with others and the innate reasons for doing this. However, as Dr Neufeld and Balick identified, the development of these social technologies, for example, have occurred faster than what human beings can biologically evolve. I put forward that the implication being that human beings are still driven to connect and relate to social technologies in the same way that they would relate to other human beings. Despite, the fact that these AI social systems do not have feelings and cannot engage with the human-users at the same level of connection as them.

It is important to explain that evolutionary theory, as identified by Lakoff and Johnson (1999), held that Darwin's evolutionary theory has broadly been misconstrued, where human beings have been viewed as "self-interest maximisers by nature" (p. 558). According to Lakoff and Johnson, while evolutionary theory was seen to be about the survival of species, including the adaption to environmental conditions, this adaptation was not in relation to 'competition', as others have metaphorized, where evolutionary accomplishment was then ascribed in social Darwinism to human reasoning. In other words, putting forward that human beings have evolved to adapt to their environment, both physically and socially, to connect and to strengthen relational bonds with others, as a means of survival.

The Theory of Embodied Relating Explained

The first two categories, '**Unconscious Feeling**' and '**Conceptualising**', consider both the visceral responses from human beings as well as the concepts that are thought in the conscious mind during the interactional experience. The two categories interrelate to form meaning and experience for human beings. This dynamic process resulted in me turning to the disciplines of neuroscience and cognitive science to more fully explicate the actions and processes involved in these categories. Cundy (2015) adeptly stated that:

> Attachment theory bridges the two realms of external and internal reality: … [Bowlby's] classical theory drew on findings from other disciplines that threw light on human experience, including Darwin's evolutionary perspective… and he incorporated these insights into psychoanalytic thinking …Bowlby formulated an understanding of the conscious and unconscious mind informed by …attachment theory [with] integrated aspects of linguistics, trauma studies, and neuroscience. (pp. xv-xvi)

Cundy (2015) posited that modern neuroscience highlighted the role of attachment in structuring brain architecture, with Cozolino (2006) maintaining that human beings "have evolved elaborate neural networks for interacting with others …These systems of

attaching, predicting, and communicating are all functions of the social brain" (p. 21). Furthermore, the work of Spinoza (Damasio, 2003) held the view that organisms (including human beings) have a natural inclination for survival and thriving, albeit unconsciously. Damasio (2003) suggested Spinoza's thinking, while strongly rooted in the biological, descended from the philosopher, Aristotle's thinking. Damasio contended that:

> understanding what feelings are, how they work, and what they mean …[for] human beings … [takes] into account advances in the social sciences, cognitive science, and biology …An understanding of the neurobiology of emotion and feelings is a key to the formulation of principles and policies capable of reducing human distress and enhancing human flourishing. (pp. 7-8)

Therefore, my research findings correlated with Damasio's writing, leading to a more in-depth understanding of human emotions and feelings, which followed multi-disciplinary perspectives, from which to enhance my theory and provide a fuller meaning of human being's interactional experiences with an AI social system. These perspectives also tie in with my first category '**Unconscious Feeling**' as a way of explaining part of the process of a human being's experience when interacting with an AI social system.

Iacoboni (2008) put forward that an extended practice in human society, suggested that language originates from more non-verbally interactive and gestural. Furthermore, Iacoboni stated that "when we humans are engaged in conversation, we tend to imitate each other's syntactical structures" (p. 88). In other words, during conversations, human beings automatically and collaboratively negotiate meaning of particular words in order to form an accurate meaning for a particular context, in a particular conversation (Iacoboni, 2008). Iacoboni (2008) also pointed out that "there are other forms of imitation and interactive alignment in a face-to-face conversation. Meaning and turn-taking are automatically negotiated; simultaneous gestures, eye orientation, and body rotations are very important for helping us make sense of what is being said" (p. 98). When Iacoboni considered how "these nonverbal forms of communication easily fall into patterns" (p. 98), he was referring to how the speaker and listener interact in non-verbal gestures and communicate by engaging in a mutual gaze into the other person's eyes, indicating the flow of the conversation non-verbally. The use of non-verbal communication; such as a gaze of the eyes, gestures and even tone of voice have been highlighted by participants when interacting with an AI social system. These tended to form part of the participants' unconscious biological and emotional responses, which human-users unknowingly seek out in others, including AI social systems. This could be defined as an embodied response.

Dualism vs. Embodiment

The viewpoint of mind-body dualism has been seen to be prevalent within Western culture (Bayer & Malone, 1998) and was reflected in Lakoff and Johnson's (1999) description of how first-generation cognitive science in the 1950s and 1960s held an Anglo-American philosophy of rigorous dualism which focused on the concepts of symbolic computation. Thus, viewing the formal functions of reason and mind as independent of the body (Haugeland, 1985), which complemented the leading theories of the 1950s and 1960s, namely, "early artificial intelligence, information-processing psychology, formal logic, generative linguistics, and early cognitive anthropology" (Lakoff & Johnson, 1999, p. 75). Lakoff and Johnson argued that the main purpose of language has been to express and communicate human beings' embodied understanding and meaning, in opposition to the classical AI view which assumed that "thoughts can all be adequately expressed in a logical language … and that reason is a matter of mechanical calculation and proceeds in a step-by-step fashion" (p. 251). Subsequently, leading to the assumption that computers had the capability to 'think rationally', and the metaphor for 'the mind as a computer', in first-generation cognitive science.

Additionally, Turkle (2011) put forward that many schools of AI, such as one branch known as 'symbolic AI', affiliates itself with a Cartesian mind/body dualism philosophy, claiming that machine intelligence can be programmed through rules and the representation of facts. However, Merleau-Ponty (1962) contended that, for human beings, 'subjects' and 'objects' were not separate elements; rather, they developed from a context, providing an integrated experience on which conceptual meanings were imposed. Even earlier, Dewey (1922, 1925) determined that the entire complex human being-environment interactions that comprise human experience is simultaneously social, cerebral, and emotional. More recently, Hubert Dreyfus and Antonio Damasio argued that human beings' bodies are quite literally apparatuses of thought; thus, holding the philosophy that the mind and body are indivisible. Dreyfus, in the 1960s, argued that in order for AI machines to be intelligent they required a body. Furthermore, the neuroscientist, Damasio, held that all thoughts and emotions are embodied. Therefore, contending that a mind/body dualism was not possible as thought and feeling cannot be split. Turkle suggested this argument could be taken to mean that "robots will never have a humanlike intelligence: they have neither bodily feelings nor feelings of emotion" (p.134).

Aristotle (384-322BC) is considered to be the key philosopher who was instrumental in regard to the theory of metaphysics as a science and the systematic investigation into the nature of Being. Current notions of metaphysics, theology, and the very essence of science itself arises from Aristotle's philosophy (Lakoff & Johnson, 1999).

I reflected on Aristotle's philosophy which maintained that forms cannot exist independent of the objects themselves and, as such, are essential to the objects and necessitate being studied in relation to them. For Aristotle, form was what phenomena are based on but embodied in a particular substance (Cohen, 2016). Aristotle's theory is consistent with Lakoff and Johnson's (1999) criterion for a theory of embodied mind, namely: "there is no such fully autonomous faculty of reason separate from and independent of bodily capacities such as perception and movement. The evidence supports, instead, an evolutionary view, in which reason uses and grows out of bodily capacities" (p. 17). As such, in the last two decades of the 20th century, a number of cognitive scientists developed a theory called the embodied mind thesis (Lakoff & Johnson, 1999), arguing that the body was vital for cognition and thinking, as opposed to the earlier symbol processing theory of only the mind.

Lakoff and Johnson (1999) recognised Dewey and Merleau-Ponty as "our greatest philosophers of the embodied mind" (p. 97), with both arguing that the mind and body are not distinct abstract entities, and that experience is embodied. Dewey argued an alternative view to the dominant conception at the time of mind-body dualism, that "the organism interacts with the world through self-guided activity that coordinates and integrates sensory and motor responses. The implication for the theory of knowledge …[that] the world is not passively perceived and thereby known" (Field, n.d.). Lakoff and Johnson referred to second-generation cognitive science as 'embodied', with human beings having an embodied understanding of meaning-making. First-generation cognitive science, however, was seen to be 'disembodied' wherein meaning was just an abstract relation amid symbols or between symbols and world affairs; and how human reasoning and understanding is linked to the body was not considered Lakoff and Johnson considered the quote that "[1]the mind is inherently embodied. [2] Thought is mostly unconscious. [3] Abstract concepts are largely metaphorical" (p. 3) to signify the three key findings of cognitive science. Thus, positing these three findings from cognitive science are discordant with key aspects of Western philosophy and suggested the comprehensive reconsideration of common current perspectives, such as Anglo-American philosophy and postmodernist philosophy (Lakoff & Johnson, 1999). Lakoff and Johnson argued that this change in how reason is understood is incongruous with the main classical philosophical perspectives, such as Cartesian dualism, that considered the "mind separate from and independent of the body" (p. 5). They refuted the notion of human beings' mind being compared to that of computer software, as human beings have 'embodied minds' whose conceptual schemes happen, are shaped by, and derive meaning via the living human body.

It would appear that while the AI social systems are becoming more humanlike, mimicking human emotions and ways of interacting, there still tends to be a dualistic philosophy from technologists who consider the rational and logical elements of AI social system development, yet fail to see, or understand, that this is not compatible with human-users' unconscious and evolutionary needs for connection and how human beings interact via embodied relating. Potentially, trigger deep, innate responses which the AI social system is unable to meet, in the same way that another human being potentially could. Iacoboni (2008) contended that the "philosophical and ideological individualistic positions especially dominant in our Western culture have made us blind to the fundamentally intersubjective nature of our own brains" (p. 152). Subsequently, these positions have been challenged by the neuroscientific analysis of the about the embodied mind. As such, Harari (2018) suggested that this individualistic position,

> may well be a chauvinistic Western fantasy, glorifying the autonomy and power of upper-class white men …evolutionary psychologists have demonstrated that most human decisions are based on emotional reactions and heuristic shortcuts rather than on rational analysis, and that while our emotions and heuristics were perhaps suitable for dealing with life in the Stone Age, they are woefully inadequate in the Silicon Age. (p. 217)

Unconscious Feeling: Thinking with Feelings

The first category of my theory, '**Embodied Relating'**, explains the visceral responses human beings have when socially interacting, pointing to humans' emotions being intrinsic to human communication, social interactions, and cognition. Throughout the interviews, I was aware of the visceral responses and feelings that participants described during their interactional experience with an AI social system. These unconscious feelings ranged from feeling creepy about the AI social system to responding in an emotionally positive way. These findings point to human emotions and feelings being fundamental to human communication, interaction, and cognition. Human beings' experiences were subjective; and the feelings of the experience stood out, as they were more about the interaction rather than the details or facts. Lakoff and Johnson (1999) reflected this unconscious process, stating that "human beings rarely pay attention to, or are even aware of, their bodily perceptions, therefore, human beings' perceptions have the illusion of being mental acts that occur independently of the unobserved body" (p. 562).

According to Damasio (2003), memory can stimulate an emotional response wherein from a neural viewpoint, images linked to a person need to be symbolised "in one or more of the brain's sensory systems, such as the visual or auditory regions" (pp. 57-58).

Furthermore, Damasio determined that emotionally competent stimuli are detected and captured speedily within the body and can bypass normal cognitive evaluation channels. In other words, undertaking self-preservation without conscious awareness of the task. As such, "when the consequences of such natural wisdom are mapped back in the brain, the result is feelings, the foundational component of our minds …[that] guide a deliberate endeavour of self-preservation" (Damasio, 2003, p. 79). These philosophies and viewpoints correlated with my findings that indicated that human beings' brains work to filter out and change information, and shape how they perceive the world. Therefore, suggesting that the brain works in a similar fashion when human beings interact with either an AI robot or avatar. For instance, overtime, human beings seem to start ignoring little details about the AI social system (how it looks, responds, and behaves) and unconsciously focus more on how they feel; that is, the core feelings experienced during the interaction. In order to further consider what unconscious feelings are and the studies around embodiment within human beings, I turned to the literature in neuroscience and cognitive science.

Neuroscience of Emotions and Feelings

In distinguishing between emotions and feelings, Damasio (2003) posited emotions to be actions and movements of the body that were visible and made public, whereas feelings were private and unseen which, he determined, were for reasons connected to biology, "emotions play out in the theatre of the body. Feelings play out in the theatre of the mind" (Damasio, 2003, p. 28). Furthermore, Damasio viewed emotions and the linked reactions underpinning them as part of the higher level, yet fundamental, processes of life regulation. Although emotions and feelings are closely linked and are processes that occur along a continuum, with emotions forming the basis for feelings, Damasio suggested that through cognitive neuroscience we can "discover how it is that we feel" (p. 28). This suggestion was warranted as a pertinent exploration of ongoing literature to support and inform my research findings. As such, Damasio determined that human beings had emotions first, where emotions played an integral role in survival and maintaining regulation and homeostasis, and feelings second because of evolution. Therefore, Damasio further posited that "there is no doubt that emotions and feelings themselves are part and parcel of what we are, personally and socially" (p. 267). Damasio (2003) described how the homeostatic process in human beings is regulated both internally and externally of the organism by building the most favourable conditions for self-preservation and effective functioning. Thus, identifying change and responding accordingly. Consequently, this homeostatic process means that:

> machinery from all the prior levels—reflexes, immune responses, metabolic balancing, pain or pleasure behaviours, drives—is incorporated in the machinery

> of the emotions-proper…, the different tiers of emotions-proper are assembled on the very same principle. ... each reaction consists of tinkered rearrangements of bits and parts of the simpler processes below. They are all aimed at the same overall goal—survival with well-being—but each of the tinkered rearrangements is secondarily aimed at a new problem whose solution is necessary for survival with well-being. The solution of each new problem is required for the overall goal to be achieved. (Damasio, 2003, p. 38)

Damasio (2003) further posited that basic emotions, as well as the conditions that cause them and the characteristic patterns of behaviour, are consistent across many cultures and are easily recognised in human beings. He theorised those social emotions, such as sympathy, embarrassment, shame, guilt, pride, jealousy, envy, gratitude, admiration, indignation, and contempt, are likely to be involved in the development of intricate cultural processes of social regulation. Furthermore, some of these human social emotional responses can occur in circumstances where one person reacts without them or the other actor being immediately aware of it. Another set of reactions, Damasio suggested, resulted from "nonconscious origin[s] shaped by learning during one's individual development" (p. 48). Consequently, putting forward the notion that both these sets of responses are "nondeliberate, nonconscious reactions—those innate and those learned" (p. 49) and might form part of the unconscious. Damasio (2003) considered emotions as a natural process, often conscious, for the brain and mind to assess the surroundings within and around the person, consequently reacting and adapting. Thus, processing both the current object but its connection to others and how it links with the past. However, Damasio also noted that:

> even when the emotional reaction occurs without conscious knowledge of the emotionally competent stimulus, the emotion signifies nonetheless the result of the organism's appraisal of the situation. Never mind that the appraisal is not made clearly known to the self …One of the main aspects of the history of human development pertains to how most objects that surround our brains become capable of triggering some form of emotion or another, weak or strong, good or bad, and can do so consciously or unconsciously. Some of these triggers are set by evolution, but some are not, instead becoming associated by our brains with emotionally competent objects by virtue of our individual experiences. (p. 55)

This statement by Damasio (2003) provides a fuller picture in relation to explaining my findings, and my subsequent theory '**Embodied Relating'**, in terms of the human-user having both conscious cognitive processes and concepts, as well as unconscious feelings that work together – in a way for human beings to adapt and interact with others, for the evolutionary purpose of survival and thriving.

Buying into the Hype and Having Expectations.

Kislev (2022) determined that "today's realities push us to recognise the benefits of these technologies" (p.16), thus buying into the hype. The outbreak of the COVID-19 pandemic and reports of extensive mental health issues as a result of social isolation, indicated that loneliness is a growing social issue that will continue to be around in the near future. Additionally, if an AI social system has been designed to look human-like in some way, human beings treat them as another social actor and expect them to have some level of social intelligence, including understanding and expressing emotions.

Embodying.

Melmed (2021) stated that "Freud's (1923) notion that the self begins as a projection of bodily sensation, plying a sort of echo chamber buildup of resonances and dissonances, …[arriving] at a notion of an embodied self that extends well beyond the individual" (p. 128). In terms of the subcategory '**embodying**', I considered Damasio's (2003) explanation that emotions are enacted in actions, such as "facial expressions, vocalizations, body postures, and specific patterns of behaviour (running, freezing, courting, or parenting)" (p. 63). Whereas a feeling, Damasio put forward "is the perception of a certain state of the body along with the perception of a certain mode of thinking and of thoughts with certain themes" (p. 86). His viewpoint was that feelings function uniquely, as their core is comprised of the thoughts that correspond with the body involved in a reactive process. Damasio stated that "whatever we feel must be based on the activity pattern of the body-sensing brain regions …The feelings we experience come courtesy of body-sensing regions" (p. 111). A term known as 'uncanny valley' was mentioned by several participants, and they spoke about this experience in a visceral and embodied way. In 1970, a Japanese roboticist, Masahiro Mori, introduced the idea of the 'uncanny valley' (Dumouchel & Damiano, 2017; Kislev, 2022). His hypothesis proposed that when robots bear a close, but not complete, resemblance to human beings, it induces a strong negative response and discomfort. Dumouchel and Damiano (2017) suggested that human beings find this "disconcerting, sometimes shocking or even repellent" (pp. 25-26), as if a living corpse. Furthermore, Kislev (2022) suggested that these robots that nearly resemble human beings provoke aversion, specifically because they do "not seem like the 'whole package' when we encounter one" (p. 104).

Gestures and Facial Expressions

Shaw (2003) put forward that "the face is obviously a means whereby we can communicate emotional states" (p. 44). Similarly, Kislev (2002) noted that when robots included gestures, human-users found them more "appealing and trustworthy…[especially] increasing [their] eye contact with the robot, smiling, and closer

physical proximity" (p. 192). This demonstrated how human-users continued to interact with AI social systems, as they would with other human beings, unconsciously feeling and conceptualising and relating in an embodied way.

Developing Trust.

Further, Damasio (2003) surmised that the brain is able to simulate certain emotional body states, through a process referred to as an 'as-if-body-loop' mechanism, which changes the emotion of sympathy into a feeling of empathy; even though individuals were unaware of the process of mirroring occurring in their own body states. I correlated Damasio's inference to the subcategory of '**developing trust**', where the human-user begins to develop trust in the AI social system as their emotion of sympathy converts to a feeling of empathy. If this process does not occur, the human-user does not develop trust in the AI social system and disengages from the interaction. Turkle (2011) considered the meaning of trust and the differences between human beings trusting other human beings versus human beings trusting AI social systems. Turkle suggested that trusting human beings takes time to develop as opposed to AI social systems where it is more about testing the system. If it responds effectively, then it is trusted.

Damasio (2003) stated that "there is growing evidence that feelings, along with …the emotions that most often cause them, play a decisive role in social behaviour" (p.140). Damasio suggested that a major trait of human behaviour that imbues civilization is the ability to think with regard to the future. Thus, human beings respond to social situations, and are shaped by social emotions, by slowly categorising the situations they experience. Therefore, whether the emotions are negative or positive they become part of human beings' social experience. This means that human beings can link categories of social knowledge (learnt or through personal experience) with the inherent, biological mechanism of social emotions and their ensuing feelings - thus further developing the human-users' trust of the AI social system. In considering human beings' vulnerabilities towards AI social systems, Turkle (2011) considered them "powerful because they invite our attachment" (p. 78). As such, Kislev (2022) contended that human beings' innate mechanisms of attachment and empathy "are much more instinctive and basic than we believe them to be" (p. 201). Additionally, Kislev suggested that as these AI social systems have become more advanced, it has contributed to drawing human-users "into an emotional trade and convince us they're worthy of our feelings" (p. 201). This emotional trade of the human-user tends to then lead them to be more complicit with the AI social system, growing more accommodating of its' technological hiccups.

Growing Accommodating.

Kislev (2022) reported that in recent research human beings who interacted with a chatbot “every day for two weeks buil[t] some kind of habit” (p. 138), which saw participants bonding with the chatbot over the 2 weeks, feeling like emotional exchange had developed that led to increased levels of attachment. Subsequently, human-users grew more accommodating to any technological issues or disruptions by the AI social system, as well as the fact that they were engaging with a machine not another human being. Turkle (2011) suggested that human-users “do whatever it takes to sustain a view of these robots as sentient and even caring” (p. 85). Referring to ‘complicity’ where the human-user is drawn into feeling in a relationship with the enlivened AI social system. According to Turkle, AI social systems “promise satisfactions, even if only in fantasy. Getting satisfaction means helping the [AI social systems], filling in where they are not yet ready, making up for their lapses. We are drawn into necessary complicities” (p. 86). The ability for human beings to alter the nature and shape of social interactions allows them to adapt to a larger range of situations (Dumouchel & Damiano, 2017).

Furthermore, Kislev (2022) cited Professor Mario Verdicchio, an expert in the ethics and philosophy of computer science at the University of Bergamo, Italy, who stated that in order for human beings to feel emotionally attached to AI social systems:

> The **suspension of disbelief** is fundamental ...I can see a direct correlation between the strength of the suspension of disbelief (i.e., the strength of my conviction that the robot is indeed entertaining feelings) and the strength of my emotional attachment to it (i.e., I can ‘love it more’, ‘have real feelings for it’). (p.139)

Thus, according to Kislev (2022), human beings interacting with AI social systems is not about enhancing their belief in the abilities of the AI social systems; rather, is more about suspending disbelief, from the human-user, that they are interacting with a machine. Therefore, the human-user is willing to believe and go along with the suspending of disbelief (Dumouchel & Damiano, 2017). This was a property that was mentioned by some participants as well.

Projection.

Laplanche and Pontalis (1988) defined projection as the “qualities, feelings, wishes or even ‘objects’, which the subject refuses to recognise or rejects in himself, are expelled from the self and located in another person or thing” (p. 349). When projection is initiated, it is experienced as closeness and a part of the ‘real relationship’. As referenced by Winnicott (1986), projection is concerned with when **the ‘other’** is perceived through the perspective of the self, leading to a modified perception of the ‘other’.

Turkle (2011) referenced psychoanalyst, Heinz Kohut, in terms of projection, suggesting that “some people may temporarily strengthen their fragile sense of self by turning another person [or AI social system] into a ‘self-object’. In the role of self-object, the other is experienced as part of the self, thus in perfect tune with the fragile individual’s inner state” (p. 7). In other words, human-users may perceive the AI social systems’ aliveness as a contender for the part of self-object (extension of oneself).

Cundy (2016) put forward that our habituated attachment styles influence our social connections, suggesting that “projections …can be overblown due to the lack of real, embodied other holding, metabolising, and mirroring our communications to counterbalance the inventions of our imaginations” (p. 196). Of note, human-users who have moved through the processes of an interaction with an AI social system, can experience what Lakoff and Johnson (1999) called “empathic projection” (p. 566); whereby they identify with the ‘experience’ of the AI social system—something that can happen with pet owners and their pets.

Anthropomorphising.

Dumouchel and Damiano (2017) explained that “anthropomorphism derives from the operation of fundamental cognitive structures, which is related to our tendency to think teleologically and to interact through dialogue” (pp. 109-110). As a result, ‘**anthropomorphic projections**’ arise in relational connections where the content of interaction is reduced; and the converser, usually an animal or an object, does not command the attributes that have been bestowed to it during the interaction. Kislev (2022) identified that human beings, while consciously aware that they are interacting with an AI social system, were still inclined to regard them as social actors – “even [being] susceptible to compliments made by machines” (p.137). Additionally, these human-users referred to the chatbot by the pronoun ‘him’, stating “it’s human to make mistakes” (p.138); thus, **anthropomorphising** the chatbot. This perception was identified in the data, where all participants referred to the AI social system as a ‘social actor’ using pronouns. Dumouchel and Damiano (2017) referred to more recent research that considered “that anthropomorphism constitutes a fundamental dimension of the human mind …[where] the propensity to attribute mental states to inanimate objects is …strongly influenced by the nature and context of the interaction” (p.109).

Getting in Flow State.

As relayed by a technologist participant, the intended outcome of AI social systems that mimic human emotions is for the interaction with the human being to natural, intuitive, and engaging. Such interaction leads to the human-user getting in a conversational and interactional flow state. Kislev (2022) further suggested that:

> By investing in freestyle conversations, developing a full human-like narrative, building a character for us, and presenting a more comprehensive personality for their [AI social systems], developers can encourage that 'suspension of disbelief' that will allow us to think of AI [social] systems as human-like companions. (p. 139)

Feeling Lulled and Forgetting.

Getting in a flow state for the human-user while interacting with an AI robot or avatar, leads to what Balick (2014) referred to the term 'magic' when considering the notion that AI social systems could seem magical in that they trigger human beings' omnipotent fantasies. Further to this, Damasio (2003) explained how an,

> emotional signal can operate entirely under the radar of consciousness. It can produce alterations in working memory, attention, and reasoning so that the decision-making process is biased toward selecting the action most likely to lead to the best possible outcome, given prior experience. The individual may not ever be cognizant of this covert operation. In these conditions we intuit a decision and enact it, speedily and efficiently, without any knowledge of the intermediate steps. (pp. 148-149)

As the connection has continued, and, perhaps strengthened further, the human-user begins feeling lulled, at an unconscious level. With the result that, "such embodied mechanisms of conceptualisation and thought are hidden from our consciousness, but they structure our experience and are constitutive of what we do consciously experience" (Lakoff & Johnson, 1999, p. 497). Leonhard (2016) considered the role of technology in inviting human beings to 'forget themselves', and what forgetting ourselves when interacting with AI social systems could mean as technology advances exponentially. Leonhard contended that "one thing is certain: technology and many of its biggest providers are doing whatever they can do to endear us to the paths of assentation and abdication, whether inadvertently or by design" (pp. 54-55). Harari (2015) noted that human beings employ a 'narrating self' that recovers memories, relays stories, and is involved with important decision-making. However, when this 'narrating self' assesses the human beings' experience "it discounts their duration and adopts the 'peak-end rule' – remembers only the peak moment and the end moment and assesses the whole experience according to their average" (pp. 343-344). As indicated in the research findings, part of interactional experience is that the human-user 'forgets' then subsequently recalls 'peak' experiential moments from their interaction with the AI social system. As such, this explanation complements the study data which determined that participants were '**feeling lulled**' and then '**forgetting**' they were interacting with an AI social system and responding as if they were interacting with another human being.

Conceptualising: Constructing an 'Other'

Dumouchel and Damiano (2017) explained that a distinguishing feature of social robots "is the ability to be perceived by those who interact with them as being present in the way a person is …a robot must give its human partner the 'feeling of being in the company of someone', of being confronted with an 'other'" (pp. 102-103). The **constructing of an 'other'** is reflected in Lakoff and Johnson's (1999) view of empathy by a human being as being the capability to espouse the viewpoint of another, which they described as "metaphorically as the capacity to project your consciousness into other people, so that that you can experience what they experience, the way they experience it" (p. 309). Lakoff and Johnson (1999) argued that in order to understand the human mind, the philosophical viewpoints of reason need to be considered. First, that reason is not disembodied; that is, reason results out of the combination of the human brain, body, and resultant bodily experience, referred to as embodiment. Second, reason is considered on the evolutionary continuum with other animals. Third, reason is a capability shared universally by all human beings. Fourth, that reason is mostly unconscious. Fifth, that reason is to a large extent metaphorical and imaginative; and sixth, that reason is connected to emotions. As such, Lakoff and Johnson argued that "reason is shaped by the body" (p. 5). Another aspect of conceptualising within the interactional experience, is how the human-user and AI robot or avatar converse and communicate with each other. In reference to human language, Lakoff and Johnson (1999) argued that language is more than 'pure syntax', as human beings' language is dynamic in nature and includes "meaning, context, perception, emotion, memory, attention, [and] action" (p. 6). Cognitive sciences were described by Lakoff and Johnson as a "scientific discipline that studies conceptual systems" (p. 10). Use of the word 'cognitive', in cognitive sciences, is to refer to "any kind of mental operation or structure that can be studied in precise terms" (Lakoff & Johnson, 1999, p. 11).

The majority of these structures and operations are found to occur at an unconscious level, meaning that visual and auditory processing, memory, thought, language and attention, as well as any unconscious inferences, all fall under the cognitive. Cognitive science also studies neural modelling of any cognitive operation, including mental metaphors, emotions, and the notion of motor operations. Lakoff and Johnson (1999) put forward that human beings have evolved to categorise as a means of survival. They noted that most categorising is automatic and unconscious, and shaped though our embodiment and experience; with a small portion of human categorisation being shaped by conscious reasoning. The concepts that human beings form are neural structures that enable them to rationally distinguish between categories and reason about them.

This reasoning is conceptually inferred using the sensorimotor system of the brain. To summarise Lakoff and Johnson (1999),

> our categories arise from the fact that we are neural beings, from the nature of our bodily capacities, from our experience interacting in the world, and from our evolved capacity for basic-level categorisation – a level at which we optimally interact with the world. (p. 30)

The above explanations from the field of cognitive science and neuroscience indicates how, through the embodied process **of conceptualising**, human beings facilitate conscious reasoning and meaning making of their experiences, actions, and situations. Through the use of words to communicate concepts, which inhabit the human mind, and obtain their meaning by way of their embodiment, Lakoff and Johnson (1999) posited that human beings, from childhood onwards, "conceptualise the world throughout embodied experiences and the shaping provided by the structure of our bodies and brains" (p. 442). As such, human beings establish the meaning of concepts via the embodied experience.

Through empirical evidence from cognitive neuroscience and embodied cognitive science, human beings' cognitive unconscious automatically influences how we understand and make meaning of our experiences, which Lakoff and Johnson (1999) referred to as "unreflective common sense" (p. 13). In other words, findings from cognitive science (Lakoff & Johnson, 1999) have determined that human reason is intrinsically linked to our bodies and brains, along with our interactions with our surroundings, which have been influenced by evolution and experience. Thus, providing a mainly unconscious rationale for our reasoning in determining what is real.

Reality

Harari (2018) highlighted that human beings have taken over the world as a result of being able to create and believe in fictional stories, yet "we are therefore particularly bad at knowing the difference between fiction and reality" (p. 306). Science fiction novels and movies have tended to portray the notion of robots as caring for human beings, which is now prevalent throughout society (Turkle, 2011). The notion of science fiction fantasies of full-fledged human-robot relationships are likely to become a reality before long. Kislev (2022) suggested that "the vision of technology satiating our emotional, intellectual, and physical needs is no longer limited to science fiction" (p. 7), with "science fiction is becoming science fact" (Leonhard, 2016, p. 10). According to Lakoff and Johnson (1999), human beings' make sense of the world based on:

> our sensory organs, our ability to move and to manipulate objects, the detailed structure if our brain, our culture, and our interactions in our environment, at the very least. What we take to be true in a situation depends on our embodied understanding of the situation, which is in turn shaped by all these factors. Truth for us, any truth that we can have access to, depends on such embodied understanding. (p. 102)

Kislev (2022) noted that "technology has already changed our expectations of reality …Every possibility is now a potential reality, and this widens our social, emotional, and sexual horizons" (p. 10). Kislev contended that as technology and embodied human beings synthesise, over the course of time, that it will impact on human beings' senses, experiences and how meaning of these experiences is perceived, thus "revolutionising what we expect from reality" (p. 11).

Filling a Gap

AI social systems could fill a gap and meet some human needs, which could include emotional, psychological, or even sexual needs. Human beings feel comfortable to share and express their emotions to AI social systems as a way for them to ponder on and work through complex occurrences and emotions. This subcategory is resonated in Kislev's (2022) findings, as he also suggested that AI social systems could fill a gap and meet a human need. The care industry for human beings (e.g., taking care of infirm, children, elderly) is one area that has received a significant amount of attention for AI social systems to fill a gap. One aspect of a worldwide loneliness issue, particularly among the elderly, is where AI social systems are being developed and used to be a companion for the human-user. The idea is that AI social systems could be more available, caring and polite that an elderly person's family or medical staff, such as nurses.

Fearing Misuse and Abuse

Impact on Human-to-Human Interactions.

Human beings have been responding to AI social systems emotionally, becoming attached to them, and even finding solace in their presence, as if they were human. Kislev (2022) echoed some of my questions around AI social systems and how they affect and change our social, emotional, and physical lives. The role of technology has change substantially from being an object aiding in connecting human beings to being the subject of affection itself. In his study, Kislev identified that people interacting with an AI social system called 'Replika', developed trust, closeness, and compassion for the AI social system. This led to the consideration that AI could substitute characteristics of human-to-human relationships and the possibility of unconventional outcomes.

These possibilities may be positive for human beings and their real relationships, or more negative where human beings' relationships with each other become ungratifying and potentially unstable, as human relationships can be complex, with messy interactions, which impacts the expectations human beings' have of each other as opposed to the more agreeable AI social systems. Consequently, Turkle (2011) considered it a blurring of the lines between AI social systems that have emotions and those that mimic emotions, as if they had them. Damasio (2003) compared living organisms and intelligent machines (such as AI social systems), emphasising that intelligent machines are "not living in the sense we are and are not likely to feel in the way we do" (p. 129).

Furthermore, Damasio highlighted that that there is no intelligent machine that can feel anything comparable to human feelings, as they have no equivalent internal life comparable to human beings. Balick (2014) considered that "there are aspects embedded within the architecture of [AI social systems] …that encourage relating as an object to objects rather than the fuller form of intersubjective relating" (pp. 101-102). Balick acknowledged that human beings' need to relate, as well as need to recognise and be recognised, has not altered; but posited that these AI social systems are impacting the central and innate humanness in how we do core human relating and how this may be shaping us.

Shifting Norms/Generational Shift

Kislev (2022) identified a clear distinction between "age, education, gender, urban/ rural divide, and the degree of exposure affect our attitude toward these new technologies" (p. 148). Thus, acknowledging that we are in the middle of a technological generational shift, with younger generations more accepting and willing to shift towards being in relationships with AI social systems. Adding to this shift in norms, as technology has improved and become cheaper, and the subsequent COVID-19 pandemic, resulted in technology industry advancing and growing significantly. Technological advancements have uncovered human needs and have contributed to fundamental changes in social norms; additionally, pointing out how the COVID-19 pandemic sped up the transition in human beings' use and adaptation to technology to combat the pandemic, connecting human-to-human online and engaging with the technology as a social actor in itself. Due to advances in technology and the expansion of the internet, human beings, especially younger generations, are keen adopters of these technologies, leading to shifts in social structure and shifting norms.

Culture.

Culture reflects how people communicate and interact with each other and how they perceive and interact with AI social systems. For example, in reflecting on culture and language in comparison between Western and Eastern cultures, historically, many Western cultures saw AI social systems as a symbol of becoming alienated and detracting from humanity. However, many Eastern cultures, such as the Japanese culture, revere technology (Dumouchel & Damiano, 2017).

Turkle (2011) stated that at the end of the 20th century, the Japanese determined that there were more aging than young Japanese population. Therefore, they decided to develop robots that would take care of the elderly—some for practical purposes and some specifically for being a companion to the elderly. This practice has become a reference point for many Americans, in the Western culture, as an example of how they could tackle taking care of their own aging population. Therefore, has Japan been at the forefront of AI robot development and use in a multitude of settings, especially in elder care. Contrary to initial assumptions, elderly people were not afraid or disconcerted with robots being used to aid their rehabilitation; rather, they seemed to find them fascinating and engaged with them naturally (Blair, 2017). Further, with regard to AI robots within the Japanese culture, it is important to consider the Japanese concept of Shinto which "holds that energetic forces called kami exist in all aspects of the world and universe. Kami infuses humans but also animals, insects, trees, rocks, and human creations, like robots or virtual artifacts" (Kislev, 2022, p. 235). Furthermore, these beliefs and concepts could rapidly become ubiquitous, especially as the world becomes more secular and shared meanings of human beings' relationships with AI social systems advance.

The Japanese government coined the terms 'Society 5.0.' (or the 5th evolution of society) in 2016, to signify the evolutionary human development in the substantial developments in robotics, biotech, AI, quantum computing, cyber physical systems, and nanotech; which, together, transform the way human beings live. Some key elements that are influenced by culture are, for example, a society's structure determining which sector(s) an AI social system might be most useful for (such as if the focus is on industry development and hospitality then robots would be designed more for that industry), and different countries would have varying priorities and funding for AI social system designers and researchers which would influence the development and output. Consequently, New Zealanders have been less likely to understand AI robots and avatars, and are not as familiar with AI social systems. Further to the aspect of culture, Krech et al. (1962) stated that "the language of a people mirrors their culture, that it can be regarded as the crystallised thought of a people" (p. 293).

Holmes (2001) postulated that human beings are bound together, both individually and as a society, by relational connections and language. Further suggesting that

> there is a grammar and syntax – structure and rules – to emotional relationships, of which we are largely unaware, yet which have a determining effect on how we feel about ourselves and …others, and which we have learned – like a language – within the microculture of family. (Holmes, 2001, p. 66)

As such, consistent with the context of qualitative research, where "language is a tool for communicating people's experiences, perspectives and practices; …language reflects reality [and] ...the factors and social processes that underpin particular phenomena" (Clarke, Braun & Hayfield, 2015, p. 227).

Bootstrapping Itself.

In terms of considering how AI might be **bootstrapping itself**, Kislev (2022) highlighted that AI social systems are becoming more advanced in their 'intellectual' capabilities, making them less contingent on human programming. Additionally, suggesting that AI social systems are becoming more sophisticated and an increased variety of functions.

Human Beings' Relationship with AI Social Systems

Cundy (2015) posited that "technology is shaping our relationship to reality and self-experience" (p. 5) which has historically been in the 'real' world. Yet, exponentially increasing, social technology relating cannot be ignored as a strong determinant in attachment experiences. It would appear that although the preeminent human drive for attachment is still present, the 'who' and 'how' has changed through the use of AI social systems. From a socio-cultural perspective, it is evident that both human beings (individually and as a society) attribute symbolic meanings to technology and that these meanings shape a co-constructed relationship between society and its technologies. Baym (2010) suggested that what a society declares about its technology is as telling of that society, as it is of the technology.

Balick (2014) posited that the relational nature of the co-constructed realities and intersubjectivity between human beings can be extended in order to consider human beings' relationships to AI social systems, and, further, to investigate how societies and cultures relate to these technologies. This relational and interactive process is what Baym (2010) termed 'social construction of technology'. Turkle (2011) put forward that "we are drawn by our humanity to give to these machines something of the consideration we give to each other. Because we reach for mutuality, we want them to care about us, as we care for them" (p. 100).

Additionally, Baym contended that human beings need to deliberate how societal circumstances foster new technologies, and that the advancement of technology is relational, in and of itself. Science fiction stories have included considering the changing nature of how technology has been perceived, namely in the service of humanity.

The introduction of AI social systems as a companion or partner, have the potential of making human-to-human relationships superfluous or even expendable. Resulting in the concern that human beings may begin to perceive each other in the same way as they do AI social systems; namely, as an object or means to gain pleasure as opposed to being in relationship as the purpose for the connecting and relating. This is resonated by Kislev (2022) contending that "AI [social systems] illicit feelings from us that are surprisingly similar to those we have in human-to-human interactions" (p. 111).

Research (Turkle, 2011; Kislev, 2022) further indicated that human-users fostered emotional attachments to their basic AI social systems, such as an AI voice assistant. Turkle (2011) suggested that "when we are confronted with an entity that [behaves in humanlike ways, such as using language and responding based on prior inputs,] our brains' default response is to unconsciously treat the entity as human" (p. 140). This was a response that was evident in the findings of my study as well.

Summary

Some science fiction has had the premise that "human consciousness is located entirely inside of our brains and that our bodies merely act as hosts for the 'self'" (Kislev, 2022, p. 178). This implicit hypothesis has, for a long time, correlated with the dominant dualistic paradigms of many advocates for this approach, with the view that AI social systems are would be able to provide all the companionship a human being would need. According to Turkle (2011), the notion that an attentive AI social system is able to fulfil the fantasy of avoiding other human relationships that are negative, even abusive, can resulting in these AI social systems coming across as "caring enough" (p. 282). Thus, human beings are willing to adapt to the technology in order to create the fantasy that there are substitutes for human beings.

Implications from the Findings

In considering the implications of this research and findings, I reflected on Leonhard's (2016) statement about the development of AI social systems that "the key emerging questions are why should it be done, who will be in charge or control it, and what it may mean for the future of humanity?" (p. 77). Furthermore, Leonhard suggested that "it is critical to determine a sustainable balance between precaution and pro-action" (p. 106), with regards to AI technology and humanity. These statements echo the underlying research implications from this study, to be discussed in the next section of this chapter.

Furthermore, the COVID-19 Pandemic has highlighted how quickly humans can transfer human-to-human interactions online, adapting to connecting online via some form of technology. There are implications in the development and use of these AI social systems interacting with human beings, something echoed by Kislev (2022) who stated that "technology is now making the transition to a relationship partner and becomes an independent actor in itself, and we must understand the implications of this change" (p. 5).

The COVID-19 pandemic has impacted how people relate and the implications this could have both at a micro and macro level. On a micro level, as technology and embodied human beings synthesise, over the course of time, this it will impact on human beings' senses, experiences and how meaning of these experiences is perceived, thus transforming human beings' perceptions and assumptions of reality. These impacts have already begun to be experienced by human-users of these social technologies and were reflected in participants' comments. At a macro level, Naughton (2012) reflected that "our society has become critically dependent on a technology that is poorly understood, not just by its users, but also by people (like government ministers) who are in a position to make decisions about how it should be regulated and controlled" (pp. 10–11). Findings in this study indicated that there was a perceived lack of governmental and regulatory bodies that understood the implications of AI social systems for human beings and society as a whole. Additionally, the lack of sufficient interdisciplinary contributions in the development and uses of AI social systems also added to the poor understanding of the impacts of AI social systems on human beings, and lack of sufficient education and transparency for the human-user and third parties that purchase this technology, for services such as customer service.

Ethics and Regulations

As such, in considering the implications of these findings, in terms of ethics and regulations, I concur with Leonhard's (2016) suggestion that "we urgently need precise yet liquid public policies; new social contracts; global digital health standards; localised, responsive regulations; and deeper responsibility and involvement of marketers and advertisers of these technologies" (p. 100). Kislev (2022) held a similar view, stating that:

> we should investigate these applications and the impact and consequently suggest ethical standards, policies, and regulations that make them safe, just, healthy, and beneficial …Especially as this progress seems inevitable, there is no way to avoid applying some guidelines to the usage (pp.176-177)

In demonstrating human beings' interactional experiences with an AI robot or avatar, through the processes demonstrated by my **'Embodied Relating'** theory, the need for

further understanding and the application of robust ethics and regulations around how this technology is developed, marketed, used by human beings and whether the 'human need' justifies the technology, as well as what protection is in place for human-users, is urgently needed. Leonhard (2016) made the comparison between the early times of nuclear weapons and many 'Silicon Valley' technologists, contending that that these "technologists seem to be proceeding as fast and as far as they can without a modicum of reflection on what issues in innovations may end up causing" (p. 121).

Of course, robust regulations would also be an imperative for AI that might 'bootstrap' itself, resulting in it being less dependent on human programming – especially AI social systems that involve interacting with human beings.

Implications for Human Beings

The impending reality concerning the continuous evolvement of technology cannot be evaded; which now includes some relational elements that mimic human beings with even more relational components being mimicked in the near future. Additionally, the implication that human-users are likely to become more and more accepting of less complex, detailed and meaningful exchanges with AI social systems, coincides with Kislev's (2022) suggestion that the quality of human conversations has worsened in recent years as a result of social media communications. Furthermore, the rate of the exponential change shaped by modern technology, creates a level of unpredictability and leads to fast-paced changes in how human beings meet our 'physical, emotional and social needs' (Kislev, 2022).

In identifying the implications of AI social systems for human beings, findings from this research are echoed in Turkle's (2011) argument that AI social systems "cheapen the notion of companionship to a baseline of 'interacting with something'. We reduce relationship and come to see this reduction as the norm" (p. 55). There was a developing sense from participants that interacting with AI social systems was inevitable and becoming more and more accepted – thus becoming the norm. Additionally, Leonhard (2016) stated that while human beings can "appear to be clumsy, complicated, slow, risky, or inefficient compared to non-biological systems, computers, and robots" (p. 23), it is nevertheless important to recognise our humanity as something that should be protected. Furthermore, while acknowledging the benefits of AI social systems, consideration and the understanding of the implications need to be given to the overuse and dependency on these AI social systems by human beings, as well as considering how human-users could be drawn into attachments with AI social systems that could alter their way of being and interacting in society.

Human beings' have evolved and adapted to their environment, both physically and socially, thus contributing to them connecting and strengthening their relational ties with others as a form of survival and thriving. The implication is considered in Dumouchel and Damiano (2017) argument that:

> Evolution having made us in such a way that we know how to recognise, by certain signs that roboticists exploit, when we are the objects of another's gaze. This ability evolved because being observed by another organism is often a biologically important situation …and sometimes a dangerous one. (p. 151)

The essence of this statement from Dumouchel and Damiano was identified within the findings of this study, further noting the implications for regulations to protect human beings from exploitation of these AI social systems, as well as the implications how human beings engage and interact with others - whether other human beings or AI robots or avatars.

Implications for Therapy

As this research was conducted with a psychotherapeutic lens, Weizenbaum's (1976) deep concern had piqued my interest as both a human researcher and psychotherapist. It has contributed to my focus on this topic, with the exponential development of AI social robots and/or digital avatars, specifically being designed for social interactions with humans, such as in mental health, healthcare, and companionship (Soul Machines, 2019, World Health Organization [WHO], 2022). Cundy (2015) questioned whether people would turn away from in-person therapy that might be slow and potentially painful, and towards online quick-fix therapy, or AI robot and avatar therapy. The implication here suggests that technologists are not yet fully understanding the techniques, processes and value of in-person psychotherapy. As well as, not understanding how human beings operate and work through psychological issues, with their therapist (Shaw, 003).

As mentioned in Initial Literature Review, Chapter 2, the first endeavour of joining AI with psychotherapy was in the 1960s and was named ELIZA. It employed Rogerian psychotherapy techniques where human-users' statements and thoughts were responded to in the form of questions. However, the systems lacked the capabilities to conduct a complete psychotherapeutic session. As developments and capabilities improved in the 1980s, AI chatbots could conduct more complex interviews to establish basic diagnoses. More recently, there are several projects aiming to create an AI virtual therapist, to help people in real-time with their mental health issues (Kislev, 2022). For example, an AI avatar has been created to help combat veterans who have been diagnosed with post-traumatic stress disorder (PTSD) (Lucas, Rizzo, Gratch, et al., 2017).

Another example is the text-based AI chatbot, Woebot, which chats with human users, concentrating predominantly on their mental health (Kislev, 2022). Kislev noted that "the fact that some users refer to Woebot as 'he' or 'she' is telling. It is getting closer to the place where the interaction with Woebot resembles that of human-to-human" (p.134). However, the concern with AI social systems as 'therapists', suggests the focus is less about the human value and meaning of life and relationships; and more about human being 'behavioural change' and emphasis on the technology and what it can do (Turkle, 2011).

Turkle (2011) viewed that a human being that talks to AI robot thought to be 'therapeutic' claiming that just by "getting something 'out' [and that] …bad feelings become less toxic when released" (p. 231). Thus, holding the assumption that a human psychotherapist is not required for the client to deal with their feeling. The concern is that this decreased expectation of what 'therapy' with an AI social system would be, results in a new acceptance of needing less from human beings and more from technology (Turkle, 2011). As such, Neufeld & Mate (2013) were of the view that AI social systems that provide psychological therapy offer "a pseudo-satisfaction for unmet attachment needs, they can be extraordinarily addictive" (p. 280). Therefore, in human-to-human psychotherapy, clients are recognised and "feel like [they] matter, having a self-image of genuine mastery can only develop in nurturing relationships with [human therapists'] care about [them]. They are the outcomes of healthy attachments" (Neufeld & Mate, 2013, p. 280). Furthermore, Holmes (2001) stated that "therapy influenced by attachment theory …it is about empathy and responsiveness, but it is also about the separateness of the therapist – his capacity to have a mind of his own …and the ways in which patients cope with the rhythm of attachment and parting that is integral to the therapeutic relationship" (p. 33).

As contended by Shaw (2003), psychotherapists include their somatic body responses as part of the therapeutic dialogue, which is co-constructed through the therapeutic engagement with the client. Something that AI social systems providing 'therapy' cannot do, yet.

Recommendations

This study has contributed new knowledge by providing a substantive grounded theory, theoretical model, and illumination to advance the understanding of human beings interactional experiences with an AI social system. Specifically, the research demonstrated that the human-users' experience involved, concurrently, unconscious thinking and conceptualising throughout the interaction.

With the influences from the fantasy of science fiction, impacting human beings' sense of reality, shifting norms and filling a gap/human need and the subsequent fear of misuse and abuse. Recommendations that developed out of this study have implications for further research, psychotherapy practice, education, ethics and AI regulatory bodies, as well as humanity in generally.

Further Research

This research examined human beings' experience interacting with an AI social system. This is a significantly understudied area and would benefit from extending the findings of this research. The AI field is extensive and while this research focused human beings' interactional experiences of both AI robots and avatars, further research into more in depth studies of human beings' interactional experiences with varying types of AI robots and avatars, such as those that look very humanlike to those that look more mechanical. This would also include looking at further research into the understanding of various aspects of AI social system embodiment for human-users. Additionally, a useful direction for further research would be to investigate the attachment styles of the human-users and how they engage with the AI social system, and what the outcome of that interaction does in terms of their attachment to the AI social system subsequently. As well as, studying how their subsequently relationships and attachment are with other human beings.

A highly valuable area for further research would be for psychotherapists to research, not only the AI social system that are marketed as 'mental health therapists' in offering 'therapy' for human beings with a range of mental health concerns including PTSD, depression and anxiety, to ensure that they are therapeutically sound, ethical and safe for the human beings that might engage with them – who may be vulnerable already. But also, to research the impacts these 'therapeutic' AI social systems have on the human beings that use them, therapeutically and ethically. Further research in a laboratory setting, observing human-AI social system interactions would add further information to examining and understanding human-user experiences interacting with AI social systems. It would also be beneficial to conduct research on a wide diversity of human-users, including a range of nationalities, ethnicities, and ages, to get a more full and complete picture of human-user experiences globally. Furthermore, an important area recommended for further research is in considering how societal situations advance new technologies, and that the progression of technology is relational, in and of itself (Baym, 2010). As well as looking at researching the impact of human connection and relationships, as a consequence of their more and more social and relational engagement with AI social systems.

Psychotherapy Practice

Within clinical practice, an important area for further research is for psychotherapists to expand their knowledge and understanding of the technological advancements that are being made in the field of AI social systems, how they are operating and impacting individuals and society, at large. In order to better understand when their clients come to them with relational, emotional or psychological issues related to the AI social system they might be interacting with, forming a 'relationship' with and becoming attached to it. It is also imperative, as part of this recommendation for further research, that as psychotherapists learn and gain more insights in the implications of AI social systems on human beings. Especially when these technologies are used as a form of providing psychological therapy, therefore it would be recommended that psychotherapists advocate for the central role they play as human psychotherapists helping human clients, and their value in working with clients towards their individual healing journeys, as compared to AI social system 'therapists'.

Education

Usually, psychotherapists in Aotearoa New Zealand – to date – have studied quite a traditional curriculum, which has not included technology in general, and Human-AI social system interactions specifically. Psychotherapy students would benefit from the knowledge and understanding of possible ways human beings engage or not with AI social system, the impact of the human-user both psychologically, socially and physically. Incorporating both the traditional psychotherapy theories, along with neuroscience and cognitive. This feels like a significant gap in the education of psychotherapists, in this day and age and would, thus benefit from further research into establishing additional education courses within the psychotherapy qualification. Due to the tendency for technologists to still hold dualistic philosophies; considering the logical elements of AI social systems, yet potentially not being aware of the unconscious and evolutionary needs of the human-user for connection and how human beings engage and are innately triggered through **embodied relating**. Furthermore, developers or technologists of these AI social systems could benefit from having interdisciplinary training, as part of their tertiary studies. For example, learning and gaining some understanding on how psychotherapy works and the value of human-to-human engagement and relationships. As well as, understanding the risks, ethics and positives to human beings as a consequence of interacting with an AI social system. These recommendations for further education were echoed in Kislev's (2022) suggestion that "educational programs dealing with [the concerns with human-tech relationships] may be much more helpful …than fighting against technological advancement" (p. 221). Thus, applying this education into a range of disciplines and fields.

Identifying the Limitations

It is essential to recognise limitations in this research which potentially impacted on the grounded theory constructed (Birks & Mills, 2015). The purpose of this study was to investigate human beings' experience interacting with an AI robot and/or avatar. Investigating what was happening and how it would inform the future.

Constraints meant that recruiting participants in Aotearoa New Zealand was challenging as there are not many opportunities for engagement with AI social systems that are open to the general public. Therefore, not generating data that included large numbers of human-users in Aotearoa New Zealand. As such, it is not known if, having more New Zealand participants would have generated comparable results. However, recruiting participants from various countries, including some from Aotearoa New Zealand, could be seen to have turned out to be a benefit in having a range of participants. Another challenge was as a large portion of people that have interacted with AI social systems did so via institutional or organizational research, the sharing of data or participants was not often possible. This created the challenge of identifying and recruiting people who met the inclusion criteria. This effected the sample size, which was the case in my study. However, through theoretical sampling of extant, elicited data and secondary data sources, I was able to reach theoretical sufficiency.

A limitation of this study, as Harari (2015) had identified in other studies, was that due to the limited access and availability of participants, data was generally gathered from participants that represented Western, Educated, Industrialised, Rich and Democratic (WEIRD) societies – which is not representative across humanity. This was unavoidable due to the limited access and availability of potential participants.

As I was peripherally aware of AI technology, specifically AI robot and avatars before engaging in this research, I needed to research and review literature on this area quite substantially, to be able to effectively interview the participants in this study. Especially as technology is changing and advancing at such speed. I acknowledge that as an emerging researcher, I was learning and expanding my understanding, research and analytical skills while conducting the research. I honed in my interviewing skills as I extended my experience and using reflexivity throughout the research process. Consistent with constructivist grounded theory, as a researcher, I acknowledged my role in the research process, where I co-constructed data with the research participants (Charmaz, 2014; Charmaz & Bryant, 2010). Potentially, my role as a researcher and psychotherapist, included both strengths and limitations to this research.

Overall, I was able to be actively interested in learning about participants' experiences with AI social systems. A limitation could have been potentially influencing the research, however, due to conducting a presupposition interview, as well as memo-writing, reflection and discussions with the Grounded Theory Group I was able to manage this. As such, I was more aware of possibly influencing the data or assuming, so responded by questioning the participants further, to deepen my understanding. I acknowledge that another grounded theory researcher may have constructed a varied theory with the identical participants.

While this research was undertaken over an extended period of time, and some of the literature I had reviewed in my initial literature no longer reflected the stage of the AI technology development, with the advent of ChatGPT 4, for example. However, the extended time did mean I was able to examine the effects of COVID-19 Pandemic on technological advancement and the human need for it, in some areas. Furthermore, as I was conducting a multi-stage literature review, I was able to generate more current data, at the ongoing and final literature review stage.

Reflecting on the Strengths of the Study

As I employed Charmaz' (2014) version of constructivist grounded theory, I considered Charmaz' criteria for evaluating grounded theory research, namely, credibility, originality, resonance, and usefulness (as outlined in Chapter 5: Methods - Stages of Analysis and Theory Construction). I revisited these criteria to evaluate the strengths of this research.

Credibility reflects the criteria of having developed a comprehensive understanding of the topic, and adequate, applicable data in order to progress with insightful questions regarding the data, while making constant comparisons and relating data generated with the detailed analysis, throughout the research process. To improve the rigour and credibility of the research, focus was on thoroughly conducting the study in line with the aim of the research, methodology, as well as the grounded theory methods that were utilised. I have shown transparency the whole way through the PhD study, in terms of my use of methodological perspective and choices regarding grounded theory processes utilised. I illuminated methodological explanations, by providing examples from data analysis, as well as demonstrating links between data, analysis and the formulating of conceptualisations in Chapters 3-5, Methodology and Methods. Using theoretical sampling, as part of the grounded theory methods enabled me to stay true data as I constructed the theory. Credibility was reinforced by the triangulation of data generation sources (interview, elicited data, and secondary data sources). Other studies about HRI primarily focused on how to make the AI social system so that human beings would engage with or specific aspects of a discipline, in relation to HRI research.

My research was unique in that it fully focused on the human beings' experience, as whole, when interacting with an AI robot or avatar. To keep the theory grounded in the data, I have included participants' voices and experiences throughout the theory and the categories which have been present in Chapter 6-9, Findings. Additionally, through the continuous use of memo-writing, reflexivity and reflection, throughout the research process, as well as supervisory consultation and participating in regular university peer-supported grounded theory group where my decisions and interpretations were examined and critiqued.

Originality relates to the importance in providing an analysis that has theoretical significance, extending current grounded theory concepts and offering new insights (Birks & Mills, 2015; Charmaz, 2014). This original grounded theory has been presented in Chapters 6-9, Findings, which outline how the theory, categories and subcategories, as well as the links between them, are thoroughly grounded in the data. As discussed, earlier in this chapter, the theory of '**Embodied Relating'** is positioned with extant literature as well as the contributions to the research. I recognise that while useful grounded theories contribute new knowledge, they additionally offer recommendations for further research – as outlined in the recommendations section.

Resonance refers to the reflected depth of the analysis and constructed theory, that represents, fully, the participants' experiences, revealing deeper meanings and insights, and providing a grounded theory that makes sense to participants and others who share their situations (Charmaz, 2014; Birks & Mills, 2015; Charmaz & Thornberg, 2020). I ensured that participants' experiences resonated in the data, theoretical concepts and subsequent theory.

Usefulness relates to the applicability of the analysis and interpretations to everyday situations, considering the generic processes as a result of the analysis and any implied consequences. Furthermore, considering whether the analysis contributes to further research in other areas and to further knowledge (Charmaz, 2014). Bryant (2017) contended that "the overarching criteria of good research should be that is makes a difference" (p. 344). Therefore, to the end, I have outlined how the findings from this research have usefulness in extending further research, as well as providing usefulness to psychotherapists, education, ethical and regulatory bodies and wider humanity, in general – as this technology is going to prevalent in all our lives in the near future.

I consider this study to be a pragmatic synthesis of Charmaz' (2014) evaluating criteria of originality and credibility which enhanced resonance, usefulness, and the ensuing value of the contribution.

Charmaz (2006) suggested that “a grounded theory that conceptualizes and conveys what is meaningful about a substantive area can make a valuable contribution” (p. 182). This theory has the potential to be used in a variety of settings and the suggested transferability into other disciplines and situations is shown. As well as the relevance, and usefulness of this theory in wider human interpersonal contexts.

Conclusion

This constructivist grounded theory study provided a theoretical explanation of human beings’ experience interacting with an AI social robot and/or digital avatar. By applying grounded theory methods, the theory of **Embodied Relating** was constructed.

An increasing number of theorists contend that the embodied mind of human beings is a fundamental part of consciousness, referred to as the ‘embodied cognition theory’ (Kislev, 2022). This suggests that the way human beings think, view and feel about themselves is linked to the way human minds and consciousness engage and experience the human body via physiological responses, language and emotion. Added to this, is the body language and non-verbal communication that contributes to and influences human beings’ connection with others. Further to Morse’s (2009) observation that “grounded theory evolved and changed – and is still changing” (p. 18), so has sociological literature evolved in regard to the body and experiences of embodiment, suggesting that embodied feelings be included as part of symbolism within interactional contexts (both in social and psychosocial-biological terms) (Waskul & Vannini, 2012). Since the philosopher Socrates, the underlying call of philosophy has been to ‘know thyself’. As such, in order to know ourselves personally, we need to understand who we are as human beings. Lakoff and Johnson (1999) stated “we are all philosophical animals” (p.551), therefore, arguing that philosophy assists us human beings to make sense of our lives and provides the opportunity for in depth insight into who we are as people and how we experience our world, as well as how we can live better lives. Kislev (2022) stated that the “technological advancements have opened the door to the convergence of humans and technology” (p. 239). This is especially true, since the onset of the COVID-19 Pandemic, which has led to the increased adoption of new types of technology, which have included applying these technologies to human relationships. This a timely new area of research, in relation Human-AI social systems interactions, from a psychotherapeutic lens. Especially when considering Dumouchel & Damiano’s (2017) statement that “social robots have the power to bring about a metamorphosis in the ecology of our social relations” (p. 169).

My theory is unique in addressing Human-AI social systems interaction from the human beings' experience. Additionally, it provides psychotherapeutic understanding of human beings' interactional experiences with AI social systems, that can be extended into other disciplines as well. I believe it is our responsibility to investigate and research these AI technologies and how they will affect human beings – how we live and relate with ourselves and each other. Especially, as there are uncertainties around what the long-term usage and engagement of these kinds of AI social systems will be for human beings.

The theory of **Embodied Relating** highlights the concurrent processes of Unconscious thinking and Conceptualising, and how these are key processes to understanding human beings' experiences and dynamic processes when interacting with an AI social system. While parts of my theory align with extant literature and extant theory, my theory as summarised as a whole, is original and contributes insights into how human beings experience **Embodied Relating** when interacting with AI social robots and/or digital avatars.

References

AI New Zealand Ltd. (2023). *What is artificial intelligence?* https://newzealand.ai/what-is-ai

Alasuutari, P. (1996). Theorizing in qualitative research: A cultural studies perspective. *Qualitative Inquiry, 2*(4), 371-384. https://doi.org/10.1177/107780049600200401

Aldiabat, K. M., & Le Navenec, C. (2011). Philosophical roots of classical grounded theory: Its foundations in symbolic interactionism. *The Qualitative Report, 16*(4), 1063-1080. https://doi.org/10.46743/2160-3715/2011.1121

Alenljung, B., & Lindblom, J. (2015). User experience of socially interactive robots: Its role and relevance. In J. Vallverdú (Ed.), *Handbook of research on synthesizing human emotion in intelligent systems and robotics* (pp. 352-364). Information Science Reference. https://doi.org/10.4018/978-1-4666-7278-9

Alvesson, M., & Sköldberg, K. (2009). *Reflexive methodology: New vistas for qualitative research* (2nd ed.). SAGE.

Anderson, D. R. (1987). *Creativity and the philosophy of C.S. Peirce*. Dordrecht: Martinus Nijhoff.

Andrews, T. (2012). What is social constructionism? *Grounded Theory Review, 11*(1), 39-46.

Asimov, I. (1950). *I, Robot*. Gnome Press.

Balick, A. (2014). *The psychodynamics of social networking: Connected-up instantaneous culture and the self*. Karnac Books.

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Sabanovic, S. (2020). *Human-robot interaction – An introduction*. Cambridge University Press.

Barney, N. (2023). *Artificial superintelligence (ASI).* https://www.techtarget.com/searchenterpriseai/definition/artificial-superintelligence-ASI

Bayer, B., & Malone, K. R. (1998). Feminism, psychology and matters of the body. In H. J. Stam (Ed.), *The body and psychology* (pp.94–199). London: Sage.

Baym, N. K. (2010). Personal connections in the digital age. Polity.

Berger, P. L., & Luckmann, T. (2011). *The social construction of reality: A treatise in the sociology of knowledge*. Open Road Media.

Biesta, G. (2010). Pragmatism and the philosophical foundation of mixed methods. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioural research* (2nd ed., pp. 95-118). SAGE. https://doi.org/10.4135/9781506335193

Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. SAGE.

Birks, M., & Mills, J. (2015). *Grounded theory: A practical guide* (2nd ed.). SAGE.

Birks, M., & Mills, J. (2019). Rendering analysis through storyline. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of current developments in grounded theory* (pp. 243-258). SAGE.

Blair, H. (2017). [Review of Mountain Mandalas: Shugendō in Kyushu, by A. G. Grapard]. *The Journal of Japanese Studies, 43*(2), 403–407. http://www.jstor.org/stable/26448199

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Prentice Hall.

Bowen, G. A. (2006). Grounded theory and sensitizing concepts. *International Journal of Qualitative Methods, 5*(3), 12-23. https://doi.org/https://doi.org/10.1177/160940690600500304

Bowlby, J. (1969). *Attachment and loss: Vol. 1: Attachment*. Hogarth Press.

Bowlby, J. (1979). *The Making and Breaking of Affectional Bonds*. London: Tavistock

Bowlby, J. (1980). Attachment and Loss. Vol. 3: *Loss, Sadness and Depression*. New York:

Bowlby, J. (1987). Defensive processes in the light of attachment theory. In D. P. Schwartz, J. L. Sacksteder, & Y. Akabane (Eds.), *Attachment and the therapeutic process: Essays in honor of Otto Allen Will, Jr., M.D* (pp. 63–79). International Universities Press, Inc.

Briggs, C. L. (2007). Anthropology, interviewing, and communicability in contemporary society. *Current Anthropology, 48*(4) 551-580. https://doi.org/10.1086/518300

Bringer, J. D., Johnston, L. H., & Brackenridge, C. H. (2004). Using computer-assisted qualitative data analysis software to develop a grounded theory project. *Field Methods, 18*(3), 245-266. https://doi.org/10.1177/1525822X06287602

Britannica, T. Editors of Encyclopaedia (2023, November 1). *Artificial Intelligence (AI): At a Glance.* Encyclopedia Britannica. https://www.britannica.com/topic/Artificial-Intelligence-AI-At-a-Glance-2235722

Broekens, J., Heerink, M., & Rosendal, H. (2009). Assistive social robots in elderly care: A review. *Gerontechnology, 8*(2), 94-103. https://doi.org/10.4017/gt.2009.08.02.002.00

Brush, K. (2021). *Chatbot*. https://www.techtarget.com/searchcustomerexperience/definition/chatbot

Bryant, A. (2002). Re-grounding grounded theory. *Journal of Information Technology Theory and Application, 4*(1), 25-42.

Bryant, A. (2009). Grounded theory and pragmatism: The curious case of Anselm Strauss. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 10*(3), Art. 2. https://doi.org/10.17169/fqs-10.3.1358

Bryant, A. (2017). *Grounded theory and grounded theorizing: Pragmatism in research practice*. Oxford University Press.

Bryant, A. (2021). Continual permutations of misunderstanding: The curious incidents of the grounded theory method. *Qualitative Inquiry, 27*(3-4), 397-411. https://doi.org/10.1177/1077800420920663

Carlino, R. (2011). *Distance psychoanalysis: The theory and practice of using communication technology in the clinic*. Karnac.

Carpenter, J. (2013). Just doesn't look right: Exploring the impact of humanoid robot integration into explosive ordnance disposal teams. In R. Luppicini (Ed.),

Handbook of research on technoself: Identity in a technological society (pp. 609-639). IGI Global.

Carpenter, J. (2016). *Culture and human-robot interaction in militarized spaces: A war story*. Routledge. https://doi.org/10.4324/9781315562698

Charmaz, K. (1990). Discovering chrome illness: Using grounded theory. *Social Science and Medicine, 30*(11), 1161-1172. https://doi.org/10.1016/0277-9536(90)90256-r

Charmaz, K. (1991). Translating graduate qualitative methods into undergraduate teaching: Intensive interviewing as a case example. *Teaching Sociology, 19*(3), 384-395. https://doi.org/10.2307/1318205

Charmaz, K. (2000). Grounded theory: Objectivist and constructivist and objectivist methods. In N. K. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). SAGE.

Charmaz, K. (2001). Grounded theory: Methodology and theory construction. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopaedia of the social & behavioural sciences* (pp. 6396-6399). Elsevier.

Charmaz, K. (2003). Grounded theory: Objectivist and constructivist methods. In N. Denzin & Y. Lincoln (Eds.), *Strategies of qualitative inquiry* (2nd ed., pp. 249-291). SAGE.

Charmaz, K. (2005). Grounded theory in the 21st century: A qualitative method for advancing social justice research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp. 507-535). SAGE.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. SAGE.

Charmaz, K. (2008a). Constructionism and the grounded theory method. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 397-412). Guildford Press.

Charmaz, K. (2008b). Grounded theory as an emergent method. In S. N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 155-172). Guildford Press.

Charmaz, K. (2009). Shifting the grounds: Constructivist grounded theory methods. In J. M. Morse, P. N. Stern, J. Corbin, B. Bowers, K. Charmaz, & A. E. Clarke (Eds.), *Developing grounded theory: The second generation* (pp. 127-154). Left Coast Press.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE.

Charmaz, K. (2017a). *Grounded theory masterclass with Emeritus Professor Kathy Charmaz*. 25-27 September 2017, University of Queensland, Australia.

Charmaz, K. (2017b). The power of constructivist grounded theory for critical inquiry. *Qualitative Inquiry, 23*(1), 34-45. https://doi.org/10.1177/1077800416657105

Charmaz, K., & Bryant, A. (2010). Grounded theory. In B. McGaw, P. P. Peterson, & E. Baker (Eds.), *The international encyclopaedia of education* (3rd ed., pp. 401-406). Elsevier.

Charmaz, K., & Mitchell, R. G. (1996). The myth of silent authorship: Self, substance, and style in ethnographic writing. *Symbolic Interaction, 19*(4), 285-302. https://doi.org/10.1525/si.1996.19.4.285

Charmaz, K., & Thornberg, R. (2020). The pursuit of quality in grounded theory. *Qualitative Research in Psychology, 18*(3), 305-327. https://doi.org/10.1080/14780887.2020.1780357

Chenitz, W. C. & Swanson, J. M. (1986). *From practice to grounded theory: Qualitative research in nursing*. Addison-Wesley.

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine, 7*. https://doi.org/10.1177/2050312118822927

Clarke, A. E. (2005). *Situational analysis*. SAGE.

Clarke, A. E., Friese, C., & Washburn, R. S. (2018). *Situational analysis: Grounded theory after the interpretive turn* (2nd ed.). SAGE.

Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic analysis. In J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (pp. 222-248). SAGE.

Cohen, S. Marc (2016). *Aristotle's metaphysics*. Stanford Encyclopaedia of Philosophy. https://plato.stanford.edu/entries/aristotle-metaphysics/#ActuPote/

Colby, K. M., Watt, J. B., & Gilbert, J. P. (1966). A computer method of psychotherapy. *Journal of Nervous and Mental Disease, 142*(2), 148-52. https://doi.org/10.1097/00005053-196602000-00005

Connor, A. O., Carpenter, B., & Coughlan, B. (2018). An exploration of key issues in the debate between classic and constructivist grounded theory. *Grounded Theory Review: An International Journal, 17*(1). https://groundedtheoryreview.com/2018/12/27/an-exploration-of-key-issues-in-the-debate-between-classic-and-constructivist-grounded-theory/

Cooley, C. H. (1902). *Human nature and social order*. Charles Scribner's Sons.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Grounded theory procedures and techniques* (3rd ed.). SAGE.

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE.

Coulson, M. C., Oskis, A., Meredith, J., & Gould, R. L. (2018). Attachment, attraction and communication in real and virtual worlds: A study of massively multiplayer online gamers. *Computers in Human Behavior*, *87*, 49-57. https://doi.org/10.1016/j.chb.2018.05.017

Cozolino, L. (2006). *The neuroscience of human relationships: Attachment and the developing social brain*. W. W. Norton & Co.

Creswell, J. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. SAGE.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*. SAGE.

Crevier, D. (1993). AI: The tumultuous search for artificial intelligence. Basic Books.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Allen & Unwin.

Crotty, M. (2003): *The foundations of social research: Meaning and perspectives in the research process* (3rd ed.). SAGE.

culture and the self. London: Karnac Books.

Cundy, L. (2015). *Love in the age of the internet: Attachment in the digital era*. Karnac.

Cundy, L. (2018). *Attachment and the Defence Against Intimacy* (1st ed.). Taylor and

Francis

CX Today. (2021). CallMiner: *Adopting AI to Help Humans Become More Humane.* https://www.cxtoday.com/contact-centre/callminer-adopting-ai-to-help-humans-become-more-humane/

Daily, S. B., James, M. T., Cherry, D., Porter, J. J. III, Darnell, S. S., Isaac, J., & Roy, T. (2017). Affective computing: Historical foundations, current applications, and future trends. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 213–231). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-801851-4.00009-4

Damasio, A. R. (2003). *Looking for Spinoza. Joy, sorrow, and the feeling brain*. Harcourt.

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B, 362*(1480), 679-704. https://doi.org/10.1098/rstb.2006.2004

Dautenhahn, K. (2013). Human-robot interaction. In M. Soegaard & R. F. Dam (Eds.), *The encyclopaedia of human-computer interaction* (2nd ed.). The Interaction Design Foundation.

Deady, R. (2011). Reading with methodological perspective bias: A journey into classic grounded theory. *The Grounded Theory Review, 10*(1), 41-57. https://groundedtheoryreview.com/2011/04/11/reading-with-methodological-perspective-bias-a-journey-into-classic-grounded-theory-2/

Denzin, N. K., & Lincoln, Y. S. (2005). *The Sage handbook of qualitative research* (3rd ed.). SAGE.

Dewey, J. (1922). Human nature and conduct: An introduction to social psychology. Henry Holt and Company. https://doi.org/10.1037/14663-000

Dewey, J. (1925). My Pedagogic Creed. *Journal of Education, 101*(18), 490-490. https://doi.org/10.1177/002205742510101803

Dewey, J. (1938a). *Experience and education*. Touchstone.

Dewey, J. (1938b). *Logic: The theory of inquiry*. Henry Holt and Co.

Dey, I. (1999). *Grounding grounded theory: Guidelines for qualitative inquiry*. Academic Press.

Digital.govt.nz. (n.d.). *Emerging Tech: Artificial Intelligence and what feeds it.* https://www.digital.govt.nz/blog/emerging-tech-artificial-intelligence-and-what-feeds-it/

Dilmegani, C. (2023). *AI Avatar: In-depth Guide for Businesses [2023].* https://research.aimultiple.com/ai-avatar/#:~:text=AI%2Dpowered%20avatars.-,What%20is%20an%20AI%20avatar%3F,Language%20Processing%20(NLP)%20algorithms.

Dictionary.com. (2023). *Artificial Intelligence*. https://www.oed.com/dictionary/artificial-intelligence_n?tl=true

Dolittle, P. E., & Camp, W. G. (1999). Constructivism: The career and technical perspective. *Journal of Vocational and Technical Education, 16*(1), 23-46. https://doi.org/https://doi.org/10.21061/jcte.v16i1.706

Dumouchel, P., and Damiano, L. (2017). *Living with Robots*. Cambridge, MA: Harvard

Dunne, C. (2011). The place of the literature review in grounded theory research. *International Journal of Social Research Methodology, 14*(2), 111-124. https://doi.org/10.1080/13645579.2010.494930

Dunne, C., & Gamze, U. B. (2020). Successfully managing the literature review and write-up process when using grounded theory methodology—A dialogue in exploration. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 21*(1), Art. 25, http://dx.doi.org/10.17169/fqs-21.1.3338

Eaves, Y. D. (2001). A synthesis technique for grounded theory data analysis. *Methodological Issues in Nursing Research, 35*(5), 654-663. https://doi.org/10.1046/j.1365-2648.2001.01897.x

Encyclopedia.com. (2019). *Human being*. https://www.encyclopedia.com/humanities/dictionaries-thesauruses-pictures-and-press-releases/human-being#:~:text=oxford,Pocket%20Dictionary%20of%20Current%20English

Engati (2021). *Artificial Narrow Intelligence*. https://www.engati.com/glossary/artificial-narrow-intelligence

euRobotics AISBL (2013). R*obotics 2020: strategic research agenda for robotics in Europe*. euRobotics Coord. Action Draft 0v42, 7th Framework Programme, European Commission, Brussels, Belgium. https://ec.europa.eu/research/industrial_technologies/pdf/robotics-ppp-roadmap_en.pdf

Fendt, J., & Sachs, W. (2008). Grounded theory method in management research: Users' perspectives. *Organizational Research Methods, 11*(3), 430-455. https://doi.org/10.1177/1094428106297812

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5), e13216. https://www.jmir.org/2019/5/e13216

Fonagy, P. (2001). *Attachment theory and psychoanalysis*. Other Press.

Frank, A. (2019). *The Rise of a New Generation of AI Avatars*. SingularityHub. Freeman & Company. https://singularityhub.com/2019/01/15/the-rise-of-a-new-generation-of-ai-avatars/

Gabatz, R. I., Schwartz, E., Milbrath, V. M., Zillmer, J. G. V., & Neves, E. T. (2017). Attachment theory, symbolic interactionism and grounded theory: Articulating reference frameworks for research. *Nursing, 26*(4), e1940017. https://doi.org/10.1590/0104-07072017001940017

Gadow, S. (1982). Body and self. A dialectic. In V. Kestenbaum (Ed.), *The humanity of the ill: Phenomenological perspectives* (pp. 86-100). University of Tennessee Press.

Geertz, C. (1973). *The interpretation of cultures*. Basic Books.

Giacomini, M. (2010). Theory matters in qualitative health research. In I. Bourgeault, R. Dingwall, & R. de Vries (Eds.), *The Sage handbook of qualitative methods in health research* (pp. 125-156). SAGE.

Glaser, B. (1998a). *Doing grounded theory. Issues and discussions*. Sociology Press.

Glaser, B. (1998b). *Gerund grounded theory: The basic social process dissertation*. Sociology Press.

Glaser, B. (2001). *The grounded theory perspective: Conceptualization contrasted with description.* Sociology Press.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory.* Sociology Press.

Glaser, B. G. (1992). *Emergence vs forcing: Basics of grounded theory analysis*. Sociology Press.

Glaser, B. G. (2005). *The grounded theory perspective III: Theoretical coding*. Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Glaser, B., & Strauss, A. (1965). Discovery of substantive theory: A basic strategy underlying qualitative research. *The American Behavioral Scientist, 8*(6), 5-12.

Glover, E. (2023). *What Is Conversational AI?* https://builtin.com/artificial-intelligence/conversational-ai

Goertzel, B. (2018, November 27). *Emotionally savvy robots: Key to a human-friendly singularity.* Retrieved from https://www.hansonrobotics.com/emotionally-savvy-robots- key-to-a-human-friendly-singularity/

Goldkuhl, G., & Cronholm, S. (2010). Adding theoretical grounding to grounded theory: Toward multi-grounded theory. *International Journal of Qualitative Methods, 9*(2), 187-205. https://doi.org/10.1177/160940691000900205

Good, I. J. (1965). Speculations concerning the first ultra intelligent machine. *Advanced in Computers*, *6*, 31-88.

Goodrich, M. A., & Schultz, A. C. (2007). Human–robot interaction: A survey. *Foundation and Trends® in Human–Computer Interaction, 1*(3), 203-275. http://dx.doi.org/10.1561/1100000005

Goulding, C. (2002). *Grounded theory: A practical guide for management, business, and market researchers*. SAGE. https://doi.org/10.4135/9781849209236

Goulding, C. (2005). Grounded theory, ethnography and phenomenology. *European Journal of Marketing, 39*(3), 294-308. https://doi.org/10.1108/03090560510581782

Grant, B. M., & Giddings, L. S. (2002). Making sense of methodologies: A paradigm framework for the novice researcher. *Contemporary Nurse, 13*(1), 10-28. https://doi.org/10.5172/conu.13.1.10

Greene, J. (2010). Knowledge accumulation: Three views on the nature and role of knowledge in social science. In W. Luttrell (Ed.), *Qualitative educational research: Readings in reflexive methodology and transformative practice* (pp. 63-77). Routledge.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29*(2), 75–91. https://doi.org/10.1007/BF02766777

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-118). SAGE.

Guba, E.G. & Lincoln, Y.S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 193-215). SAGE.

Harari, Y. N. (2015). *Sapiens : a brief history of humankind*. New York :Harper.

Harari, Y. N. (2017). *Homo deus: A brief history of tomorrow*. Vintage Publishing.

Harari, Y. N. (2018). *21 lessons for the 21st century*. Vintage Books, New York.

Hashemi-Pour, C. (2023). *Conversational AI*. https://www.techtarget.com/searchenterpriseai/definition/conversational-AI

Haugeland, John (1985). *Artificial Intelligence*: The Very Idea. Cambridge: MIT Press.

Heath, H., & Cowley, S. (2004). Developing a grounded theory approach: A comparison of Glaser and Strauss. *International Journal of Nursing Studies, 41*(2), 141-150. https://doi.org/10.1016/S0020-7489(03)00113-5

Herman-Kinney, N. J., & Verschaeve, J. M. (2003). Methods of symbolic interactionism. In L. T. Reynolds & N. J. Herman-Kinney (Eds.), *Handbook of symbolic interactionism* (pp. 213-252). AltaMira Press.

Hickman, L. A. (2009). John Dewey: His life and work. In L. A. Hickman, S. Neubert, & K. Reich (Eds.), *John Dewey between pragmatism and constructivism* (pp. 3-18). Fordham University Press.

His Holiness Tenzin Gyatso, The 14th Dalai Lama. (2012). *A Human Approach to World Peace*. London: Wisdom Publications.

Hoare, K. J., Mills, J., & Francis, K. (2012). Dancing with data: An example of acquiring theoretical sensitivity in a grounded theory study. *International Journal of Nursing Practice, 18*(3), 240-245. https://doi.org/10.1111/j.1440-172X.2012.02038.x

Holliday, A. (2002). *Doing and writing qualitative research*. SAGE.

Holloway, I. (2008). *AZ of qualitative research in nursing and healthcare*. Wiley-Blackwell.

Holmes, J. (2001). *The search for the secure base: Attachment theory and psychotherapy*. Brunner-Routledge.

Hood, J. (2007). Orthodoxy versus power: The defining traits of grounded theory. In A. Bryant & K. Charmaz (Eds.), *Handbook of grounded theory* (pp. 151-164). SAGE.

Howell, K. (2013). *An introduction to the philosophy of methodology*. SAGE.

Hunter, J. L., & Krantz, S. (2010). Constructivism in cultural competence education. *Journal of Nursing Education, 49*(4), 207-214. https://doi.org/10.3928/01484834-20100115-06

Hutchison, A. J., Johnston, L. H., & Breckon, J. D. (2010). Using QSR-NVivo to facilitate the development of a grounded theory project: An account of a worked example, *International Journal of Social Research Methodology, 13*(4), 283-302. https://doi.org/10.1080/13645570902996301

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology, 60,* 653-670. https://www.annualreviews.org/doi/10.1146/annurev.psych.60.110707.163604

Iacoboni, M. (2009). *Mirroring people: the science of empathy and how we connect with others*. 1st Picador ed. New York, N.Y., Picador.

IBM. (n.d.). *Deep blue*. https://www.ibm.com/history/deep-blue

Jeon, Y. (2004). The application of grounded theory and symbolic interactionism. *Scandinavian Journal of Caring Sciences, 18*(3), 249-256. https://doi.org/10.1111/j.1471-6712.2004.00287.x

Kanda, T., & Ishiguro, H. (2013). *Human-robot interaction in social robotics* (1st ed). CRC Press.

Kelle, U. (1995). Theories as heuristic tools in qualitative research. In I. Maso, P. Atkinson, S. Delamont & J. Verhoeven (Eds.), *Openness in research: The tension between self and other* (pp. 33-50). van Gorcum.

Kelle, U. (2005). Emergence vs. forcing of empirical data? A crucial problem of grounded theory reconsidered. *Forum: Qualitative Social Research, 6*(2), Art. 27: https://doi.org/10.17169/fqs-6.2.467

Kelle, U. (2007). The development of categories: Different approaches in grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 191-213). SAGE.

Kennedy, B., & Thornberg, R. (2018). Deduction, induction, and abduction. In U. Flick (Ed.), *The Sage handbook of qualitative data collection* (pp. 49-64). SAGE.

Kenny, M., & Fourie, R. (2015). Contrasting classic, straussian and constructivist grounded theory: Methodological and philosophical conflicts. *The Qualitative Report, 20*(8), 1270-1289. https://doi.org/10.46743/2160-3715/2015.2251

Kislev, Elyakim. (2022). *Relationships 5: How AI, VR, and Robots Will Reshape Our Emotional Lives*. New York, NY: Oxford University Press.

Konecki, K. T. (2011). Visual grounded theory: A methodological outline and examples from empirical work. *Sociological Review, 41*(2), 131-160. https://doi.org/10.5613/rzs.41.2.1

Krech, D., Crutchfield, R. S., & Ballachey, E. L. (1962). *Individual in society: A textbook of social psychology*. McGraw-Hill.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

Kusow, A. M. (2003). Beyond indigenous authenticity: Reflections on the insider/outsider debate in immigration research. *Symbolic Interaction, 26*(4), 591-599. https://doi.org/10.1525/si.2003.26.4.591

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books

Laplanche, J., & Pontalis, J. B. (1988). *The language of psychoanalysis*. London. Karnac Books.

Law, T., Chita-Tegmark, M., Rabb, N., & Scheutz, M. (2022). Examining attachment to robots: Benefits, challenges, and alternatives. *ACM Transactions on Human-Robot Interactions, 11*(4), Art. 36. https://doi.org/10.1145/3526105

LeCun, Y., Bengio, Y., & Hinton, G. (2015) Deep learning. *Nature 52*1, 436-444. https://doi.org/10.1038/nature14539

Lempert, L. B. (2007). Asking questions of the data: Memo writing in the grounded theory tradition. In A. Bryant & C. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 245-264). SAGE.

Leonhard, G. (2016). Technology Vs. Humanity: The Coming Clash Between Man and Machine. The future agency. Zurich. Switzerland.

Liebenberg, L., Didkowsky, N., & Ungar, M. (2012). Analysing image-based data using grounded theory: The Negotiating Resilience Project. *Visual Studies, 27*(1), 59-74. https://doi.org/10.1080/1472586X.2012.642958

Lincoln, Y.S. and Guba, E.G. (1985) Naturalistic Inquiry. SAGE, Thousand Oaks, 289-331. http://dx.doi.org/10.1016/0147-1767(85)90062-8

Locke, K. (2007). Rational control and irrational fere-play: Dual thinking modes as necessary tension in grounded theorising. In A. Bryant & K. Charmaz (Eds.), *Handbook of grounded theory* (pp. 565-579). SAGE.

Lohr, S. (2016, October 17). *IBM is counting on its bet on Watson and paying big money for it. New York Times*. Retrieved from https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html?emc=edit_th_20161017&nl=todaysheadlines&nlid=62816440

Long, M., Verbeke, W., Ein-Dor, T., & Vrtička, P. (2020). A functional neuro-anatomical model of human attachment (NAMA): Insights from first- and second-person social neuroscience. *Cortex, 126,* 281-321. https://doi.org/10.1016/j.cortex.2020.01.010

Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus, 52*(2), 131-139. https://doi.org/10.1080/00380237.2018.1544514

Lucas, G.M., Rizzo, A.A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. (2017). Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Frontiers Robotics AI*, 4, 51. https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2017.00051

Mackenzie, D. (1993). *Inventing accuracy: A historical sociology of nuclear missile guidance*. MIT Press Academic.

Mallon, R. (2019). Naturalistic approaches to social construction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2019 ed.). Metaphysics Research Lab, Stanford University.

McCarthy, J., Minsky, M., Rochester, N. & Shannon, C. (1955, August 31). *A proposal for the Dartmouth Summer Research Project on artificial intelligence*. Retrieved from http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

McCorduck, P. (2004). *Machines who think* (2nd ed.). A. K. Peters, Ltd.

Mead, G.H. (1934). *Mind, Self, and Society from the Standpoint of a Social Behaviorist*. University of Chicago Press: Chicago.

Mead, G.H. (1938). *Stages in the act: Preliminary statement, Essay 1 in the philosophy of the act* (C. W. Morris Ed.). University of Chicago.

Melmed, M. L. (2021). Bound by infinities: Technology, immediacy and our environmental crisis. *Ata: Journal of Psychotherapy Aotearoa New Zealand 25*(1), 123–132.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge & Kegan Paul.

Mills, J., Bonner, A., & Francis, K. (2006). Adopting a constructivist approach to grounded theory: Implications for research design. *International Journal of Nursing Practice, 12*(1), 8-13. https://doi.org/10.1111/j.1440-172X.2006.00543.x

Minsky, M. (2006). *The emotion machine: Commons sense thinking, artificial intelligence, and the future of the human mind*. Simon & Schuster.

Mitchell, S. A. & Black, M. (1995). *Freud and beyond: A history of modern psychoanalytic thought.* Basic Books.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Higher Education, New York.

Moravec, H. (1988). *Mind children: The future of robot and human intelligence*. Harvard University Press.

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research, 5*(2), 147-149. https://doi.org/10.1177/104973239500500201

Morse, J. M. (2007). Sampling in grounded theory. In A. Bryant & C. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 229-244). SAGE.

Mulhall, A. (2003). In the field: Notes on observation in qualitative research. *Journal of Advanced Nursing, 41*(3), 306-313. https://doi.org/10.1046/j.1365-2648.2003.02514.x

Naughton, J. (2012). *From Gutenberg to Zuckerberg; What you really need to know about the internet*. London: Quercus.

Neufeld (2022). *Hold on to your kids doctor Gordon Neufeld*. ABTY podcast YouTube always better than yesterday. https://www.google.com/search?q=hold+onto+your+kids+neufeld&rlz=1C1EJFC_enNZ910NZ910&oq=hold+onto+your+kids+neufeld&aqs=chrome..69i57.8887j0j7&sourceid=chrome&ie=UTF-8#fpstate=ive&vld=cid:2816c602,vid:mzp5XC5lYa0

Neufeld, G. (2015). *Raising children in the digital world*. Workshop held November, 12, 2014 in Auckland, New Zealand.

Neufeld, G., & Mate, G. (2013). *Hold on to your kids: Why parents need to matter more than peers.* Random House.

Oktay, J. (2012). *Grounded theory*. Oxford University Press.

Olsen, S. (18 August 2006). *Spying an intelligent search engine*. CNET. https://www.cnet.com/tech/services-and-software/spying-an-intelligent-search-engine/

Peirce, C. S. (1931-1958). *Collected papers of Charles Sanders Peirce* (vols. 1–8; P. Weiss, C. Hartshorne, & A. W. Burks, Eds.). Harvard University Press.

Peirce, C. S. (1960). *Collected papers of Charles Sanders Peirce. Vol. I: Principles of philosophy; Vol. II: Elements of logic* (A. W. Burks, Ed.). Harvard University Press.

Peirce, C. S. (1979) *Collected papers of Charles Sanders Peirce. Vol. VII: Science and philosophy* (A. W. Burks, Ed.). Harvard University Press.

Piaget, J. (1964). Part 1: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching, 2*(3), 176-186. https://doi.org/10.1002/tea.3660020306

Picard, R. W. (1997). *Affective computing*. MIT Press.

Pidgeon, N., & Henwood, K. (1997). Grounded theory: Practical implementation. In J. T. E. Richardson (Ed.), *Handbook of qualitative research methods for psychology and the social sciences* (pp. 86-101). The British Psychological Society.

Priya, K. R. (2010). The research relationship as a facilitator of remoralization and self-growth: Post earthquake suffering and healing. *Qualitative Health Research, 20*(4), 479-495. https://doi.org/10.1177/1049732309360419

Priya, K. R. (2019). Literature review in grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of current development in grounded theory* (pp. 392-411). SAGE.

Rabb, N., Law, T., Chita-Tegmark, M., & Scheutz, M. (2021). An attachment framework for human-robot interaction. *International Journal of Social Robotics*, *14*, 1-21. https://doi.org/10.1007/s12369-021-00802-9

Radii. (2018). *Future of robots at 2050 with Sophia and David Chen of Hanson Robotics. Future of X forum.* Hangzhou's 2050 Conference. Retrieved from https://radiichina.com/radii-presents-future-of-x-at-2050/radii-presents-future-of-robots-at-2050-with-sophia-and-david-chen-of-hanson-robotics/

Ramalho, R., Adams, P., Huggard, P., & Hoare, K. (2015). Literature review and constructivist grounded theory methodology. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 16*(3), Art. 19. https://doi.org/10.17169/fqs-16.3.2313

Ramanathan, T. (2023). *Natural language processing*. Encyclopedia Britannica. https://www.britannica.com/technology/natural-language-processing-computer-science

Reichertz, J. (2007). Abduction: The logic of discovery of grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 214-228). SAGE.

Richardson, R., & Kramer, E. H. (2006). Abduction as the type of inference that characterizes the development of a grounded theory. *Qualitative Research, 6*(4), 497-513. https://doi.org/10.1177/1468794106068019

Russell, S. J., & Norvig, P. (2003). *Artificial intelligence: A modern approach*. (2nd ed.). Prentice Hall.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach, global edition.* Pearson Education.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). SAGE.

Salvini, A. (2019). The methodological convergences between symbolic interactionism and constructivist grounded theory. *Review of Qualitative Sociology, 15*(3), 10-29. http://dx.doi.org/10.18778/1733-8069.15.3.02.

Schön! Magazine. (2021). *Schön! presents | replika ai: your new friend*. https://schonmagazine.com/replika-film/

Schurz, G. (2008). Patterns of abduction. *Synthese, 164*(2), 201-234. https://doi.org/10.1007/s11229-007-9223-4

Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118-137). SAGE.

Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 189-213). SAGE.

Schwandt, T. A. (2003). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In N. Denzin & Y. Lincoln (Eds.), *The landscape of qualitative research* (pp. 292-331). SAGE.

Sciutti, A., Mara, M., Tagliasco, V., & Sandini, G. (2018). *Humanizing human-robot interaction*. IEEE. https://doi.org/10.1109/MTS.2018.2795095

Shaw, R. (2003). *The embodied psychotherapist: The therapist's body story*. Hove, UK: Brunner/Routledge. https://doi.org/10.4324/9780203420812

Shen, Y., Guo, D., Long, F., Mateos, L. A., Ding, H., Xiu, Z., Hellman, R. B., King, A., Chen, S., Zhang, C., & Tan, H. (2020). Robots Under COVID-19 Pandemic: *A Comprehensive Survey. IEEE access : practical innovations, open solutions*, 9, 1590–1615. https://doi.org/10.1109/ACCESS.2020.3045792

Smythe, L. (2012). Discerning which qualitative approach fits best. *New Zealand College of Midwives Journal, 46*, 5-12.

Soul machines. (2019). *Soul Machines*. https://www.soulmachines.com/

Stanfield II, J. H. (2011). *Black reflective sociology: Epistemology, theory, and methodology* (1st ed.). Routledge. https://doi.org/10.4324/9781315432892

Stern, P. N., & Porr, C. J. (2011). *Essentials of accessible grounded theory*. Left Coast Press.

Stonier, T. (1992). The evolution of machine intelligence. *In Beyond information*. Springer. https://doi.org/10.1007/978-1-4471-1835-0_6

Strang, E. (2017). Soul Machines unveils its first emotionally intelligent, lifelike avatar. *Idealog*. Retrieved from https://idealog.co.nz/tech/2017/02/soul-machines-unveils-its-first-emotionally-intelligent-lifelike-avatar

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. SAGE.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). SAGE.

Strübing, J. (2007). Research as pragmatic problem-solving: The pragmatist roots of empirically-grounding theorizing. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 580-601). SAGE.

Suddaby, R. (2006). What grounded theory is not. *Academy of Management Journal, 49*(4), 633-642. https://doi.org/10.5465/amj.2006.22083020

Tavory, I., & Timmermans, S. (2019). Abductive analysis and grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of current developments in grounded theory* (pp. 532-546). SAGE.

Techtarget. (2018-2019). (AI) Artificial Intelligence. Contributor(s): Ed Burns and Nicole Laskowski. https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence

Thistoll, T., Hooper, V., & Pauleen, D. (2016). Acquiring and developing theoretical sensitivity through undertaking a grounded preliminary literature review. *Quality & Quantity: International Journal of Methodology*, *50*(2), 619-636. https://doi.org/10.1007/s11135-015-0167-3

Thornberg, R. (2012). Informed grounded theory. *Scandinavian Journal of Educational Research, 56*(3), 243-259. https://doi.org/10.1080/00313831.2011.581686

Thornberg, R., & Charmaz, K. (2012). Grounded theory. In S. Lapan, M. Quartaroli, & F. Riemer (Eds.), *Qualitative research: An introduction to methods and designs* (pp. 41-67). Jossey-Bass.

Thornberg, R., & Charmaz, K. (2014). Grounded theory and theoretical coding. In U. Flick (Ed.), *The Sage handbook of qualitative analysis* (pp. 153-169). SAGE.

Thornberg, R., & Dunne, C. (2019). Literature review in grounded theory. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of current development in grounded theory* (pp. 206-221). SAGE.

Turing, A. M. (October 1950). Computing machinery and intelligence. *Mind, 49,* 433-460.

Turkle, S. (2011). *Alone together*. New York, NY: Basic Books.

Urquhart, C. (2007). The evolving nature of grounded theory method: The case of the information systems discipline. In A. Bryant & K. Charmaz (Eds.), *The Sage handbook of grounded theory* (pp. 339-360). SAGE.

van den Hoonaard, W. C. (1997). *Working with sensitizing concepts: Analytical field research*. SAGE.

Van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Althouse Press.

Vassilopoulos, A. P., & Georgopoulos, E. F. (2010, January 1). 5 - *Novel computational methods for fatig life modeling of composite materials* (A. P. Vassilopoulos, Ed.). ScienceDirect; Woodhead Publishing. https://www.sciencedirect.com/science/article/abs/pii/B978184569525550005X

Uneeq. (2023). *Loneliness and the role of conversational AI companions.* https://www.digitalhumans.com/blog/loneliness-and-the-role-of-conversational-ai-companions

Viljoen, B. (2017). *Exploring the impact of Millennials' use of social media on their attachment experiences, and its possible implications for psychotherapy – a thematic analysis* [Master's thesis, AUT University]. https://hdl.handle.net/10292/10634

Vygotsky, L. (1978). Interaction between learning and development. In M. Cole, S. John-Steiner, S. Scriber, & E. Souberman (Eds.), *Mind and society* (pp. 79-91). Harvard University Press.

Walton, D. (2004). *Abductive reasoning*. The University of Alabama Press.

Ward, K., Hoare, K. J., & Gott, M. (2015). Evolving from a positivist to constructionist epistemology while using grounded theory: Reflections of a novice researcher. *Journal of Research in Nursing, 20*(6), 449-462. https://doi.org/10.1177/1744987115597731

Waskul, Dennis & Vannini, P. (2012). *Body/Embodiment: Symbolic Interaction and the Sociology of the Body*. Ashgate Publishing, Ltd.

Weizenbaum, J. (1976*). Computer power and human reason*. W.H. Freeman & Company.

World Health Organization [WHO]. (2023). Meet Florence 2.0, she can give you advice on a healthier lifestyle and mental health. https://www.who.int/campaigns/Florence

Winnicott, D. W. (1986). The theory of the parent-infant relationship. In P. Buckley (Ed.), *Essential papers on object relations* (pp. 233–253). New York University Press.

You, S., & Robert, L. P. (2018). Emotional attachment, performance, and viability in teams collaborating with Embodied Physical Action (EPA) robots. *Journal of the Association for Information Systems*, *19*(5), 377-407. https://doi.org/10.17705/1jais.00496

Young, R. A., & Collin, A. (2004). Introduction: Constructivism and social constructionism in the career field. *Journal of Vocational Behavior, 64*(3), 373-388. https://doi.org/10.1016/j.jvb.2003.12.005

Zendesk. (2023). *Deep learning vs. machine learning*. https://www.zendesk.com/blog/machine-learning-and-deep-learning/

Appendices

Appendix A: AUTEC Application Approval

Auckland University of Technology Ethics Committee (AUTEC)

Auckland University of Technology
D-88, Private Bag 92006, Auckland 1142, NZ
T: +64 9 921 9999 ext. 8316
E: ethics@aut.ac.nz
www.aut.ac.nz/researchethics

24 October 2019

Brian Rodgers
Faculty of Health and Environmental Sciences

Dear Brian

Ethics Application: 19/390 **Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future**

I wish to advise you that a subcommittee of the Auckland University of Technology Ethics Committee (AUTEC) has **approved** your ethics application.

This approval is for three years, expiring 22 October 2022.

Non-Standard Conditions of Approval

1. Please ensure that the consent form is received by the participant for viewing at the same time as the Information Sheet, rather than just before the interview.

Non-standard conditions must be completed before commencing your study. Non-standard conditions do not need to be submitted to or reviewed by AUTEC before commencing your study.

Standard Conditions of Approval

1. The research is to be undertaken in accordance with the Auckland University of Technology Code of Conduct for Research and as approved by AUTEC in this application.
2. A progress report is due annually on the anniversary of the approval date, using the EA2 form.
3. A final report is due at the expiration of the approval period, or, upon completion of project, using the EA3 form.
4. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form.
5. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
6. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.
7. It is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard.

AUTEC grants ethical approval only. You are responsible for obtaining management approval for access for your research from any institution or organisation at which your research is being conducted. When the research is undertaken outside New Zealand, you need to meet all ethical, legal, and locality obligations or requirements for those jurisdictions.

Please quote the application number and title on all future correspondence related to this project.

For any enquiries please contact ethics@aut.ac.nz. The forms mentioned above are available online through http://www.aut.ac.nz/research/researchethics

Yours sincerely,

Kate O'Connor
Executive Manager
Auckland University of Technology Ethics Committee

Cc: brigitte.viljoen@aut.ac.nz; Keith Tudor

AUTEC Application Approval Amended

Auckland University of Technology Ethics Committee (AUTEC)

Auckland University of Technology
D-88, Private Bag 92006, Auckland 1142, NZ
T: +64 9 921 9999 ext. 8316
E: ethics@aut.ac.nz
www.aut.ac.nz/researchethics

30 March 2021

Elizabeth Day
Faculty of Health and Environmental Sciences

Dear Elizabeth

Re: Ethics Application: **19/390 Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future**

Thank you for your request for approval of amendments to your ethics application.

The amendment to the data collection protocol has been approved.

Non-Standard Conditions of Approval

1. Amendment of the Information Sheets as follows:
    a. Update the name of the Executive Secretary of AUTEC to Dr Carina Meares;
    b. In the privacy section replace 'AUT Cloud storage' with 'AUT Network Drive';
2. Addition of the following bullet point in the Consent Forms, selecting the appropriate language for each one:
    a. I understand that the blog or vlog/post will be used for academic purposes only and will not be published in any form outside of this project without my written permission.

Non-standard conditions must be completed before commencing your study. Non-standard conditions do not need to reviewed by AUTEC before commencing your study but please forward updated documents for our file.

I remind you of the **Standard Conditions of Approval**.

1. The research is to be undertaken in accordance with the Auckland University of Technology Code of Conduct for Research and as approved by AUTEC in this application.
2. A progress report is due annually on the anniversary of the approval date, using the EA2 form.
3. A final report is due at the expiration of the approval period, or, upon completion of project, using the EA3 form.
4. Any amendments to the project must be approved by AUTEC prior to being implemented. Amendments can be requested using the EA2 form.
5. Any serious or unexpected adverse events must be reported to AUTEC Secretariat as a matter of priority.
6. Any unforeseen events that might affect continued ethical acceptability of the project should also be reported to the AUTEC Secretariat as a matter of priority.
7. It is your responsibility to ensure that the spelling and grammar of documents being provided to participants or external organisations is of a high standard.

AUTEC grants ethical approval only. You are responsible for obtaining management approval for access for your research from any institution or organisation at which your research is being conducted. When the research is undertaken outside New Zealand, you need to meet all ethical, legal, and locality obligations or requirements for those jurisdictions.

Please quote the application number and title on all future correspondence related to this project.

For any enquiries please contact ethics@aut.ac.nz. The forms mentioned above are available online through http://www.aut.ac.nz/research/researchethics

(This is a computer-generated letter for which no signature is required)

The AUTEC Secretariat
Auckland University of Technology Ethics Committee

Cc: brigitte.viljoen@aut.ac.nz

Appendix B: Participant Information Sheet

Participant Information Sheet

People who have experienced interacting with Artificial Intelligence (AI) social robots and/or avatars.

Date Information Sheet Produced:

17 October 2019

Project Title

Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future.

An Invitation

Kia ora, my name is Brigitte Viljoen and I am conducting my PhD at AUT. I would like to invite you to participate in my research. This research will contribute to the completion of my PhD. Please note that participation in this research is voluntary and that you may withdraw at any time prior to the completion of data collection (5 December 2020. If you feel there may be potential conflict of interest issues, please note that your choice to participate or not will neither advantage nor disadvantage you.

What is the purpose of this research?

The purpose of this research is to add to the body of knowledge regarding human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future. The primary focus is to benefit both developers and users of this technology, and the general wider community. This research would involve me (the researcher) conducting an individual semi-structured interview with you (participant) as you would have experienced interacting with an AI social robot and/or avatar. The main aim of the semi-structured interviews is to ask you about your subjective experience of your interaction with AI social robot/avatar.

This research is part of my PhD qualification, and I am required to successfully complete my thesis within a set time limit. Therefore, the purpose of this research is to develop an explanatory theory about the social interactions and experiences of the participants, such as yourself. Through the process of analysis the aim would be to build a succinct conceptual theory that explains processes and actions in relation to the research question, with the intention that the theory generated and the 'process' described, in this research, may be transferable to other contexts, such as within in business, education, health care and innovation. The findings of this research may be used for academic publications and presentations.

How was I identified and why am I being invited to participate in this research?

You have been identified to participate in this research by responding to my advertisement looking for participants, and by contacting me directly to express your interest in participating. You have been invited to participate in my research as you have met the inclusion criteria of having experienced interacting with Artificial Intelligence (AI) social robots and/or avatars.

How do I agree to participate in this research?

You would first contact me to express your interest in participating, via email, after seeing my advertisement for participants. When you contact me, your email address and phone number will be collected by me and we would discuss the inclusion criteria during an initial telephone conversation (no more than 10 minutes). At this time it will be decided whether you would like to continue to the interview process. Participants will be accepted on a first-come basis.

I would give you (as a participant) a consent form before commencing the interview, which would need to be agreed to, signed and returned to me before the interview could commence. The interview whether face-to-face or video-call interview will be no longer than 60 minutes.

Your participation in this research is voluntary (it is your choice) and whether or not you choose to participate will neither advantage nor disadvantage you. You are able to withdraw from the study at any time. If you choose to withdraw from the study, then you will be offered the choice between having any data that is identifiable as belonging to you removed or, allowing it to continue to be used. However, once the findings have been produced, removal of your data may not be possible. The research design encourages mutual respect and benefit by ensuring your participation is voluntary, and by providing to you with comprehensive information about the research and the nature of participation.

What will happen in this research?

The interview process will be built on semi-structured questions and will take no longer than 60 minutes. The principle role of participants (such as yourself) is that of sharing information and providing data. You will be asked to share your perspectives of how you experienced your interaction with an AI avatar and/or social robot. Interview questions will be semi-structured so as to enable you a large degree of freedom in how you respond, ensuring that you can feel able to participate and be heard in a way that is meaningful to you. You would be invited to check the data (your interview transcript) and approve it.
The data collected will only be used for the purposes for which it has been collected.

What are the discomforts and risks?

You are likely to experience minimal discomfort or embarrassment when discussing the research topic.

How will these discomforts and risks be alleviated?

To help alleviate any potential feeling you may have of discomfort you will have the confidentiality procedures explained to you, as well as, data handling and reporting of findings. In addition, you have the right to discontinue the interview at any time. I have completed a Masters training in psychotherapy, and will be able to monitor your level of discomfort as seen in manner, tone and non-verbal cues. This will allow me to attend to your experience of being heard, respected and not judged during the interview process.

What are the benefits?

The potential benefits of this research to you (as a participant) could be gaining further insight into your own experiences with AI social robots and/or avatars. As this research is part of my PhD qualification, the benefits to me (the researcher) are significant of this will assist me in answering my research question. The benefits of this research to the wider community could be potentially significant, as it is probable that AI social robots and avatars will continue to develop for ever-growing applications.

How will my privacy be protected?

You would be actively protected from deceit, harm and coercion and would be participating on a voluntary basis, ensuring consent is informed and by informing you that you have the right to withdraw consent at any time without negative consequence. You will be reminded that you may stop the interview, or recording of the interview, at any time. Privacy is protected and any potentially identifying material will be disguised or excluded. My training in psychotherapy will enable me to be open about and aware of a potential power imbalance in the interview process. This openness will include discussions with you about my needs (as the researcher) and my genuine interest in your experience and perspectives. Diversity will be respected through open questioning in an enquiring manner, facilitating full and frank responses.
You would also be invited to offer feedback on the interview process. Any identifying demographics with participant identification numbers will be stored separately from the research data, as will signed consent forms. The latter will be held in secured AUT Cloud storage by the primary supervisor with any original paper copies shredded. I intend to transcribe the interviews personally, and the names of the participants will be protected and they would be given pseudonyms. Should there be a necessity to pay a transcriber, a confidentiality agreement would first need to be signed by the transcriber. Additionally, the interview process would take place in a private space that would protect the confidentiality of the participants. The information that is documented either physically or electronically will be kept secure either in a locked cupboard or a password protected computer.

What are the costs of participating in this research?

There are not likely to be any direct costs. However, the interview would require 60 minutes of your time.

What opportunity do I have to consider this invitation?

You have one month to consider this invitation, after which time you would receive one follow-up email or phone call should I not have heard back from you by this timeframe.

Will I receive feedback on the results of this research?

Yes, as you would have kindly provided me with information for my research, I would provide a summary of findings for your information by emailing you.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Dr Elizabeth Day, email: Elizabeth.day@aut.ac.nz, and a work phone number: +64 9 921 9999 or Mobile: +64 204211000.

Concerns regarding the conduct of the research should be notified to the Executive Secretary of AUTEC, Kate O'Connor, *ethics@aut.ac.nz* , 921 9999 ext. 6038.

Whom do I contact for further information about this research?

Please keep this Information Sheet and a copy of the Consent Form for your future reference. You are also able to contact the research team as follows:

Researcher Contact Details:

Brigitte Viljoen, email: brigitte.viljoen@aut.ac.nz or on mobile number 0204 057 0804.

Project Supervisor Contact Details:

Dr Elizabeth Day, email: Elizabeth.day@aut.ac.nz, and a work phone number: +64 9 921 9999 or Mobile: +64 204211000.

Approved by the Auckland University of Technology Ethics Committee on *24th October 2019*, AUTEC Reference number *19/390*.

Appendix C: Advertised Flyer

AUT
TE WĀNANGA ARONUI O TĀMAKI MAKAU RAU

Have you had an experience interacting with Artificial Intelligence Social Robots and/or Avatars?

Kia ora I am Brigitte Viljoen,

I would like to invite you to participate in my research. This research will contribute to the completion of my PhD requirements at AUT. The title of my thesis is:

Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future.

- The role of participants is to be interviewed, to gain more information, in order to attempt to address the research question.
- Privacy is protected and any potentially identifying material will be excluded.
- A 60-minute interview would take place, face-to-face where possible or via video calling services such as Zoom. Interview would be semi-structured questions in order to explore the lived experience of the participant who has interacted with an Artificial Intelligence Social Robot and/or Avatar.
- Participants will be selected on a first-come basis, who meet the inclusion criteria.
- An intended outcome of the study would be to provide both further understanding about this topic and to build a theory that explains processes and actions in relation to the research question, with the anticipation that the theory generated in this research may be useful to other contexts.
- **If you are interested and would like to discuss further, please contact me (Brigitte Viljoen) by email: brigitte.viljoen@aut.ac.nz or on mobile number 0204 057 0804.**

Thank you.

Approved by the Auckland University of Technology Ethics Committee on 24th October 2019 AUTEC Reference number 19/390 Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future

Note: The Participant should retain a copy of this form.

Appendix D: Participant Interview Consent Form

Consent Form

Project title: ***Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future.***

Project Supervisor: ***Dr Elizabeth Day***

Researcher: ***Brigitte Viljoen***

- O I have read and understood the information provided about this research project in the Information Sheet dated 17/10/2019.
- O I have had an opportunity to ask questions and to have them answered.
- O I understand that notes will be taken during the interviews and that they will also be audio-taped and transcribed.
- O I understand that taking part in this study is voluntary (my choice) and that I may withdraw from the study at any time without being disadvantaged in any way.
- O I understand that if I withdraw from the study then I will be offered the choice between having any data that is identifiable as belonging to me removed or allowing it to continue to be used. However, once the findings have been produced, removal of my data may not be possible.
- O I agree to take part in this research.
- O I wish to receive a summary of the research findings (please tick one): Yes O No O

Participant's signature: .....

Participant's name: .....

Participant's Contact Details (if appropriate):

.....

.....

.....

.....

Date:

Approved by the Auckland University of Technology Ethics Committee on 24th October 2019 AUTEC Reference number 19/390

Note: The Participant should retain a copy of this form.

Appendix E: Researcher Safety Protocol

All interviews for this project will be completed by [name of researcher]. These guidelines will be followed to ensure [name of researcher] is safe when interviewing in people's homes.

Advance preparation

- Gather background information. This can be clinical risk assessment information or prior telephone contact with the proposed participant. If using telephone contact it is useful to ask about current health issues, who else may be present during the interview (including pets), and directions to the participant's home. This establishes a level of rapport and allows some insights into participants' conversational style and emotional regulation.
- Schedule the interview for as early in the day as possible and always within office hours and in daylight.
- Check out the location where the interview will take place either through local knowledge (researcher or informant) and familiarise with location, entry, and exit from the area.
- Ensure that the vehicle to be used is in good working order, has sufficient fuel, and has no visible items on display prior to the visit.
- Carry the minimum equipment necessary to the interview and the minimum amount of cash.
- Dress appropriately for the research setting; flat shoes, trousers, smart casual.
- Ensure your mobile phone is charged fully and has emergency contact numbers programmed in. Keep the phone switched on at all times.

Research team

- Leave full details of the location and time of the participant contact with a member of the research team who is able to respond to emergencies.
- Phone the research team member just before the participant contact and phone again when it is over and the researcher and accompanier have reached a safe area.
- Give the estimated time of the interview and ask the research team member to phone the researcher's mobile when the allocated time has passed.
- Agree on a code sentence that if said to the research team member by phone will trigger an immediate call for the police to attend the participant contact site. An example, 'Mary you will have to cancel my meeting with Martha Carney today'.

Participant contact

- Park the vehicle as close to the participant's home as possible ensuring it is parked so that the exit can be prompt.
- Identify the safe exits from the participant's home as you go in.
- Conduct the interview in a public room where possible.
- Provide no personal details to the participant beyond the researcher name and contact number provided on the participant information sheet.

Adapted from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964461/

Appendix F: Example Script and Interview Questions from Interview 1

Scripted interview memory aids

First call memory aid:

Hello [NAME], My name is Brigitte Viljoen. If you recall, I am the PhD candidate who invited you to take part in my AUT University PhD research about your experience(s) interacting with an artificial intelligence avatar and/or social robot.

Is now a convenient time for this call to check your details and to arrange a time for a Zoom or face-to-face interview? Please can you complete the consent form that I have emailed to you and have it returned to me via email.

Are there any questions you need answered about this study before we carry on? ..... (Have info sheet handy & go through consent form questions /points...) Thank you for that.

Now, to explain a little about the process, I will be interviewing a number of people for this study. This means that I need to make a time to talk to you in a few days/ weeks' time. I'll need to talk with you for about an hour about your experience(s) interacting with an artificial intelligence avatar and/or social robot.

What would be a convenient day of the week and time of the day to do that? So let's plan a date now? ... Would you like me to text or e-mail [date] and [time] to you for your diary? *That's great*. So I shall be in touch with you the day before to confirm the interview time, and we'll go from there. If you have any questions in the meantime, please feel free to give me a call. Thank you very much for your time today. Bye. OR This means that I need to make a time to talk to you in about a month's time. Are you happy to wait that long? ...

OR

... time to call back to arrange ... No problem. So I shall be in touch with you in a couple of weeks to arrange the interview time, and we'll go from there. If you have any questions in the meantime, please feel free to give me a call. Thank you very much for your time today. Good bye.

Day Before Confirmation: Hello [NAME], My name is Brigitte Viljoen. If you recall, I am the PhD candidate who invited you to take part in my AUT University PhD research about your experience(s) interacting with an artificial intelligence avatar and/or social robot. Is this still OK? It will be for around an hour – are you still happy with that? And just to remind you that I'll be recording our chat but that I'll write our talk up afterwards so I can listen to you properly... Is that still OK? Great, now are there any questions you need answered before tomorrow? ..... I'm looking forward to talking tomorrow. Thank you, good bye.

The interview memory aid: Hello [NAME], this is Brigitte. Are you comfortable and ready for our interview? Are there any questions you'd like to ask first or are you happy to carry on? Great – So just to remind you, I am the PhD candidate who invited you to take part in my AUT University PhD research about your experience(s) interacting with an artificial intelligence avatar and/or social robot. I can confirm I am in a private space so no one else can hear you or know who I am talking to.

So I have questions to ask to help us get started talking about your experience(s) interacting with AI avatars and/or social robots – there are no right or wrong answers to anything I ask – and don't be afraid to ask me to explain a question if I don't make sense. But why don't you begin by telling me about ……

Initial Questions:

1. How would you define and explain an Artificial Intelligent (AI) social robot and an Artificial Intelligent (AI) avatar?
2. Please can you tell me about your interactional experience with an AI social robot or avatar?
3. As you look back on your experience interacting with AI social robots and avatars, are there any other events/aspects that stand out in your mind? Could you describe [each one] it? How did this event/aspect effect what happened? How did you respond to _____ [the event; resulting situation]?
4. Did you notice, if at all, what your first experience was with this AI _____? [If so,] what was it like? What did you think then? How did you happen to _____? Who, if anyone, influenced you in anyway? Tell me about that.
5. What is your general view of people in the world, how they are towards others and towards you?
6. How would you define and explain an Artificial Intelligent (AI) social robot and an Artificial Intelligent (AI) avatar?
7. Please can you tell me how you came to have an interactional experience with an AI social robot or avatar?
8. Did you notice, if at all, what your first experience was with this AI _____? [If so,] what was it like? What did you think then? How did you happen to _____? Who, if anyone, influenced you in anyway? Tell me about that.
9. Could you describe the events that led up [or preceded] to this experience?
10. What contributed to _____?
11. What was going on in your life then? How would you describe how you view AI social robots/avatars before your experience with them? How, if at all, has your view of your experience changed?
12. How would you described the person you were then?

PROMPTS:

That's interesting…

Could you explain that a little more…

Let's see, you said… Just how do you mean that…

How do you feel about ……?

Do you mind telling me …

Well, in general what would you say…

Taking everything into consideration what do you think?

On the basis of the way things look to you now what do you think?

How did you feel about the machine when you first saw it? How do you feel about it now?

Intermediate Questions:

1. What, if anything, did you know about AI social robots and avatars?
2. Tell me about your thoughts and feelings when you learned about AI social robots and avatars?
3. What happened next?
4. Who, if anyone, was involved? When was that? How were they involved?
5. Tell me about how you learned to interact with AI social robots and avatars.
6. How, if at all, have your thoughts and feelings about AI social robots and avatars changed since your experience with them?
7. What positive changes, if any, have occurred in your life since your interaction with AI social robots and avatars?
8. What negative changes, if any, have occurred in your life since your interaction with AI social robots and avatars?
9. As you look back on your experience interacting with AI social robots and avatars, are there any other events/aspects that stand out in your mind? Could you describe [each one] it? How did this event/aspect effect what happened? How did you respond to _____ [the event; resulting situation]?
10. Could you describe the most important lessons you learned through your experience interacting with AI social robots and avatars?
11. How would you describe your online interactions, for example on social media? Do you find you do more or less interacting with others online via social media, than in 'real life'? How long have you been using social media? From what age?
12. Do you have a fear of AI social robots and avatars? If so, what is your fear based on and in what context?
13. What is your general view of people in the world, how they are towards others and towards you?
14. Did you feel you grew up in a safe environment with supportive caregivers? How have they contributed to who you are today?

Ending Questions:

1. What do you think are the most important/valuable ways, if any, to incorporate AI social robots and avatars into society? How come?
2. Tell me about how your views may have changed since you have experienced interacting with AI social robots and avatars?
3. How has this experience affected you?
4. After having this experience, what advice would you give someone who was going to have an interaction with an AI social robots and avatars?
5. Is there anything you might not have thought about before that occurred to you during the interview?
6. Is there anything else you think I should know to understand your experience with AI social robots and avatars better?
7. Is there anything you would like to ask me?

Closing the interview:

The interview time is now been reached. Thank you very much for your time today – I appreciate your willingness to talk to me… If in time you wish to add something you are very welcome to call me back – my mobile is on the information sheet you have. Also, can I ask whether you are happy for me to call you back if there is anything I need to check with you about what you've said to me today? *YES / NO*

I will be interviewing more people within the course of the year. It will take about 12 months for me to put together what you and the other participants have said into a report. Once I have collected and analysed all the data, I would be happy to email a summary of the findings, for your information. Once I have done that, shall I send a copy to you? Have you any final questions?

Thank you again.

Appendix G: Transcriber Confidentiality Agreement

Confidentiality Agreement

For someone transcribing data, e.g., audiotapes of interviews.

Project title: ***Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future.***

Project Supervisor: ***Dr Elizabeth Day***

Researcher: ***Brigitte Viljoen***

- O I understand that all the material I will be asked to transcribe is confidential.
- O I understand that the contents of the tapes or recordings can only be discussed with the researchers.
- O I will not keep any copies of the transcripts nor allow third parties access to them.

Transcriber's signature: .....

Transcriber's name: .....

Transcriber's Contact Details (if appropriate):

.....

.....

.....

.....

Date:

Project Supervisor's Contact Details (if appropriate):

.....

.....

.....

.....

Approved by the Auckland University of Technology Ethics Committee on 24th October 2019 AUTEC Reference number 19/390

Note: The Transcriber should retain a copy of this form.

Appendix H: Researcher Observation Protocol

Researcher Observation Protocol

An observation protocol is a document that identifies how observation will be conducted as part of research and covers the following points:

- How people will be recruited;
- How people will be informed about the observation;
- How people will consent to the observation;
- What will be observed and what data will be collected;
- How the data will be collected;
- How any deception involved will be managed;
- The data collection instrument.

Observation may often be very intrusive on people's privacy and the observation protocol should help the researcher to ensure that their observation is sufficiently focussed so as to minimise this intrusion and to adequately provide the data required to achieve the research's aims. As each research study is quite focussed and unique, there is no standard template that can be applied across disciplines. It is more important to create a protocol that meets the needs of your study in relation to the above areas of focus.

AUTEC Application Number and Project Title

19/390 Exploring human beings' experience with AI avatars and/or social robots in an interactional context, and how will it inform our future

Applicant

Dr Elizabeth Day

Primary Researcher

Brigitte Viljoen

How will people be recruited and observed

The researcher will use the inclusion/exclusion criteria to search for extant data, such as from online social media (such as blogs and videos) that is published in a public forum and documentation such as published research that is directly related to the research question and serves to address the research question. As a result, for this extant data people will not be directly recruited.

How will people be informed and give consent about the observation

As the extant data would be already on public forums and the observation would be documented anonymously in the research, there would be no requirement to inform the people being observed and they have already consented to being observed via social media. Even so, where possible the researcher would seek to find contact details of those being observed to advise them of the research intentions and seek consent.

What will be observed and what data will be collected

In line with the research question, the behaviours, reactions and language and nonverbal communication, and general information about the people in a particular environment that display the experiences of their interaction with an AI social robot and/ or avatar would be observed.

Data would be collected by note-taking and/or transcribing the interaction, as well as, collecting extant data from other sources, such published writing (i.e. social media blogs and/or other specific research data).

"Grounded theory [research] represents both a method of inquiry and a resultant product of that inquiry" (Tie, Birks & Francis, 2019, p. 3) and encompasses the thorough use of particular methods and processes. As such the data would be collected with an open mind, using the research question and previous data collected and coded as a guide in identifying key criteria in the subsequent data collection (theoretical sampling). Birks and Mills (2015) define theoretical sampling as "the process of identifying and pursuing clues that arise during analysis in a grounded theory study" (p. 68).

Coding and Analysis

Birks and Mills (2019) suggest that the GT research methods are fluid and interactive, where the methods and processes are interconnected and are a repeat of actions and interactions that describe and inform the GT research process. According to Charmaz (2006), central to the GT method is the requirement of using constant comparative analysis techniques, as well as, the researchers' engagement. Through making constant comparisons, interacting and interpreting the data, codes and categories, the researchers' theoretical understanding would develop as these analytic features of the research categories are defined that subsequently lead to these features being rigorously analyzed, as they emerge.

- Through the memoing process, the researcher's contemplations would inform their decisions made about sampling, coding, categories and theoretical relationships within the analysis, providing additional explanations about the GT findings.
- Coding forms part of the analytical process by ascertaining repetitive theoretical ideas and similarities, and can be classified within repeated phases; namely, in constructivist grounded theorists such as Charmaz, initial, focused and theoretical coding.
- Theoretical sampling would allow leads to be identified and followed from the analysis and progress to sampling new participants or material that provides relevant information, thus developing a theoretical category.
- Data analysis would commence with the researcher being open and curious about the data, and through the simultaneous process of analysis and data collection and writing memos to document analytic thinking and methodological assumptions in relation to developing theory and data analysis.
- As the analysis takes a clear theoretical stance, through how data is coded and categorised, theory generation begins to take shape around the nature of relationships and experiences being studied (Charmaz, 2006). As stated by Lingard, Albert & Levinson (2008) "the theory emerges through a close and careful analysis of the data" (p.459).

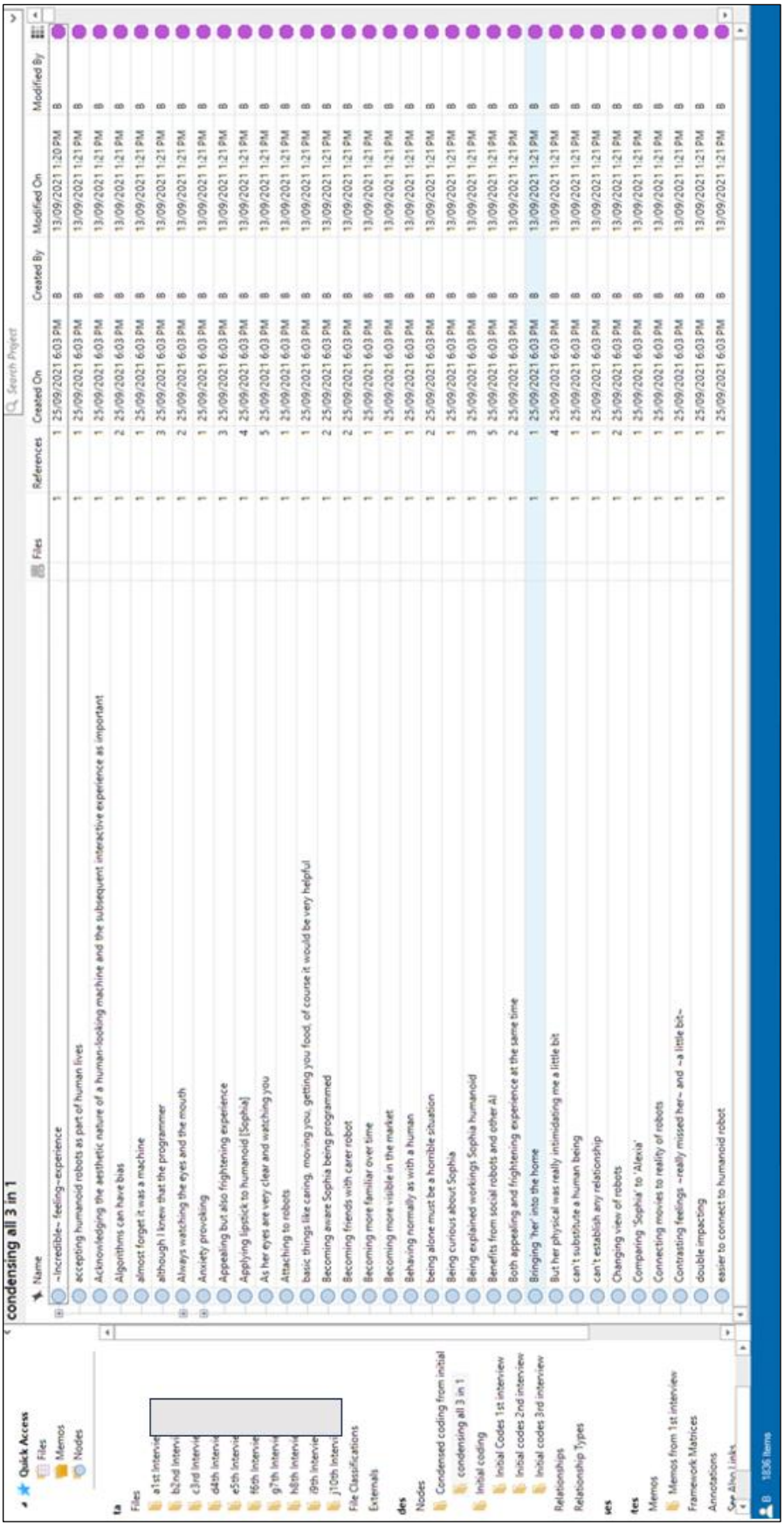
The GT analysis of this proposed study would be shaped by the theoretical content and direction of the study which may influence the amount or type of data collections methods and mean following numerous positions of inquiry, while applying rigour throughout the research process. Therefore, constructing "...an *interpretive rendering* of the worlds we study rather than an external reporting of events and statements" (Charmaz, 2006, p. 184).

Research question:

How do human beings experience artificial intelligence (AI) social robots and/or digital avatars in an interactional context? What is happening and how will it inform the future?

Description of Observation	
Type of Observation	
Description	
Data collection method (circle all relevant methods)	• Transcribing • note-taking /memoing • writing from social media • writing from published sources
Possible criteria looking out for from previous coding and comparative analysis	Understanding what an artificial intelligence social robot means?
	First noticing in the AI social robot interaction?
	Explore how the participate may be looking at the robot's eyes/face?
	Does the participant seem to be comfortable and/or relaxed with the AI social robot? Do they seem to feel safe? Participant reactions to the AI social robot?
	Did it seem like the participant felt heard by the AI social robot?
	Did it seem like there was any fear about/or the AI social robot?
	Did it seem like there was a sense of uncanny valley?

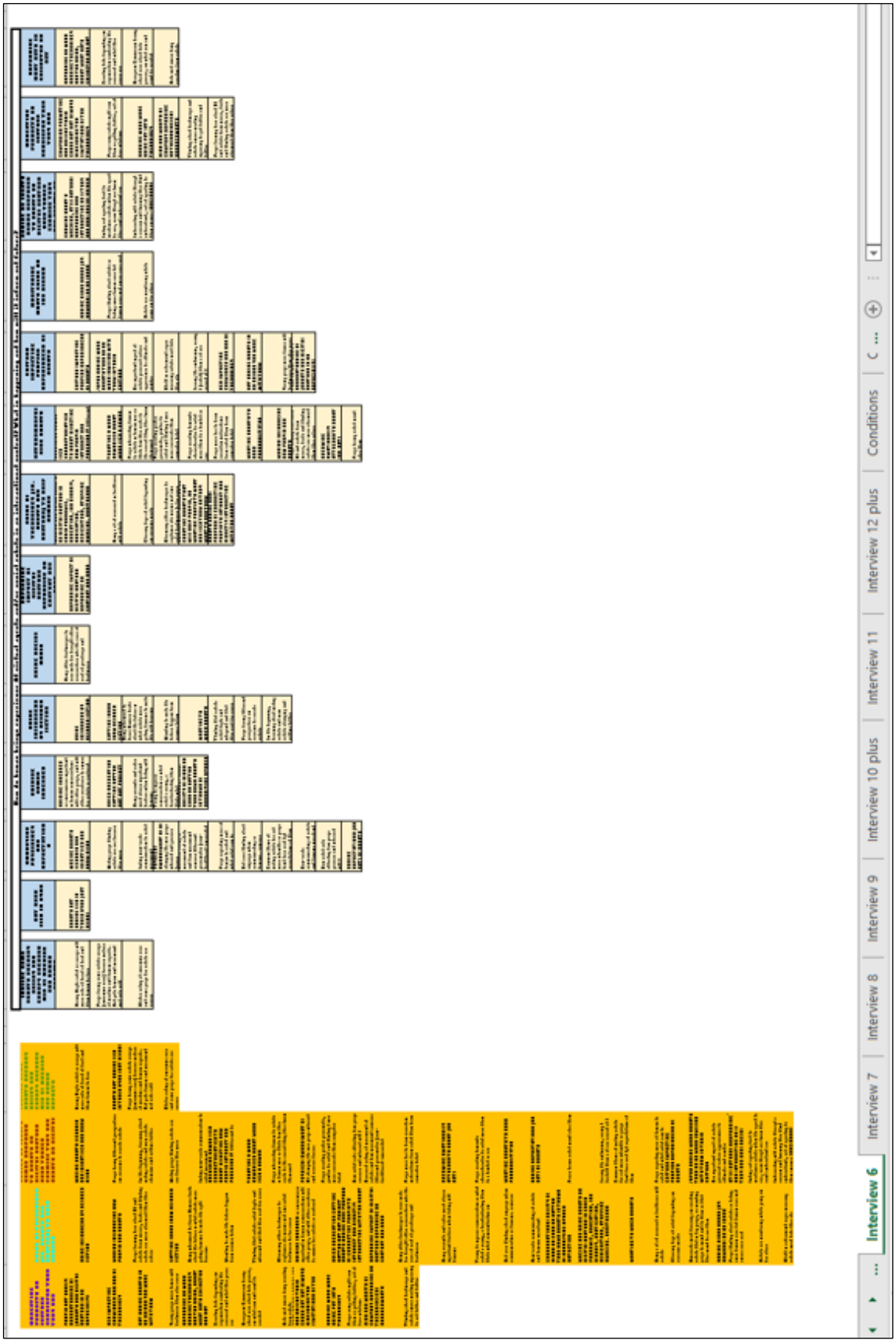
Appendix I: Example of Identified Incidents from Interview 1-3



Appendix J: Example of Interview Data and Initial Coding in Excel

B	C	D	E
Participant 13:	our first kind of customer into police national headquarters was somebody who read it online and wanted to be the first person to use it. So he came really early and waited outside our doors, to be the first person to have used Ella so there is I guess sort of expectation and inquisitiveness and um wanting to be part of kind of what the new, what the new world kind of looks like. Um if that answers your questions um.	After media launch, person waited early outside doors to be first person to have used AI social system	From media launch, people waiting to be first to use DH
		Having expectation and inquistiveness and wanting to be part of what new world looks like	Having expectation and inquistiveness and wanting to be part of what new world looks like
Interviewer:	Yeah thank you. Was there any sort of like do you know the term uncanny valley?	Not knowing meaning of uncanny valley	Not knowing meaning of uncanny valley
Participant 13:	Not really.		
Interviewer:	So um they often refer to AI robots that look very human or digital humans like sometimes there's this uncanny valley feeling of um, you know it's human but it's not human it feels weird like it's a bit of an uncomfortable feeling.		
Participant 13:	Yeah, yeah it's an unusual feeling I think for some because their brain you know from a psychological perspective she looks human you know and the quality of the art work is all intents and purposes she, she is real but she's not real and so your, your brain gives psychologically you want to try and talk to her in a, in a human way and I, because I've been in the build process I was very um okay with that by the time she got released ah but for somebody that meets her for the first time there's this nervousness you know like do I um you know do I look at her in the eye, do I, do I um, do I raise my voice, can she hear me?! Um do I speak you know quite succinct, you know using, you know trying to mouth every word, will she be able to un, understand me, can even some people tried to reach out and touch her, it's like still a computer screen, still one dimension	Some having unusal feeling [uncanny valley]	Some having unusal feeling [uncanny valley]
		Experiencing (of humans' brain) from a psychological perspective AI social system looks human and (the quality of the art work is all intents and purposes) she is real but she's not real and so your brain gives psychologically you want to try and talk to her in a human way	Psychologically experiencing digital human as real
			Psychologically human brain wanting to talk with digital human in human way
		Being in the build process, so ok with real/not real experience and knowing by time digital human released what people might be feeling when having first experience	Getting a sense of what people's fist time experience might be with digital human
		Meeting AI social system for first time, there's nervousness and wondering if need to make eye contact or raise voice for Ai social system to hear, speak succintly	Feeling nervous and uncertain about meeting digital human for first time and how to engage
		Looking digital human in the eyes	Looking digital human in the eyes
		Some people trying to reach out and touch her (even though it is a computer screen and one dimensional)	Reaching out and wanting to touch DH (even though on a screen)
Interviewer:	That is interesting		
Participant 13:	But that's how people you know the, the human part of their brain was like she's real and I just want to see if her hair is real but actually you know	People with human part in their brain 'seeing' digital human as real and wanting to see if her hair is real but can't touch it [through the screen]	Human brain seeing DH as real
Interviewer:	You can't touch it yeah.		Wanting to touch DH hair (but can't through screen)
Participant 13:	You can't touch it. Yeah but then the rational side of their brain was like it's still a computer screen it's like I don't, I don't get it. So there's this conflict, this conflict and I think that's where it's not so much un, uncomfortable it's just unusual I think.	Rational side of human brain knowing it's still a computer scree	Rational side of human brain knowing it's still a computer screen
		"I don't get it" - Conflicting experiences between feeling and rational, feeling unsual more than uncomfortable	Conflicting experiences(between feeling and rational)

Appendix K: Example of Focused Coding in Excel



Appendix L: Example of Follow-up Questions to Interview 10

Below are some follow-up questions, it would be great if you have time over the next 2-3 weeks to respond to as many of these as you can:

Researcher:
When you spoke about user projection. Please can you elaborate…?

Participant 10:
… if you engage with people, they will find their own sense in things which is what I mean by user projection.

Researcher:
Do you think AI conversation system should have 'emotions'?

Participant 10:
Personally, I do, but I feel that the ability for that to presented to the people using the system should be with their consent. I'm not much of a one for people like rights being given to AI systems.

Researcher:
How come people wanted you to put avatars on the front end of the conversational AI?

Participant 10:
Back 20 years ago I think it was about making a computer screen more approachable I suppose. We'd only relatively recently got past a command line and the terminology was still stuff like "cyberspace" …Right now what is and isn't deepfake and where the uncanny valley is heading has come up, but this is more academic that commercial.

Researcher:
With regards to [an avatar] there was a reference to the ability to build relationships. Please can you elaborate and explain.

Participant 10:
You'd need to experience (or at the very least see) the show to understand how people get absorbed into the conversation. Even cynical people forget where they are and what they are doing, somewhere between suspending disbelief that a system could have such a sophisticated conversation and enjoying the feeling of being in a relationship with an "other" in the anthropological sense of the term. Being excited to be scared kind of stuff.

Researcher:
Please can you elaborate on why and how building trust with conversational AI (in various forms) is valued and how this trust can be built?

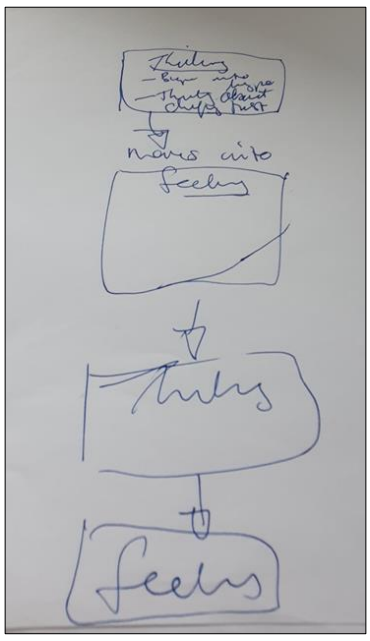
Participant 10:
Right now from an ethical/governance point of view the importance of "trust" is that people allow their data points to be used, which isn't necessarily a good thing. From the point of view of a system that is going to "help" someone to do something trusting that it is doing the right thing, whether it is a transaction or a healthcare process, is also a requirement. But how trust is built must resonate with how people trust one another I would think.

Appendix M: Memos

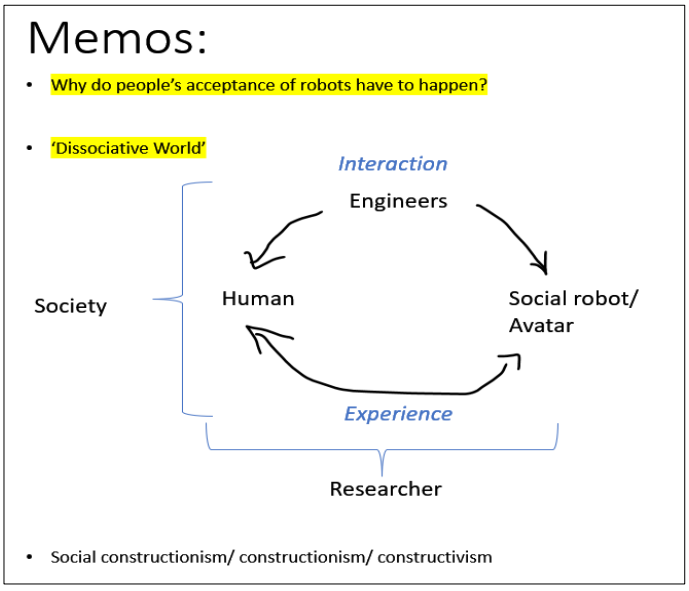
Photograph of Spider Web 15/2/2019

Something about the spider's web resonated with my PhD research question and what I was discovering. Therefore, I needed to capture this in an image and have had it stuck on my PC monitor through the PhD process.

(25/8/22)



(8/12/21)



Appendix N: Memo

(15/1/2020)

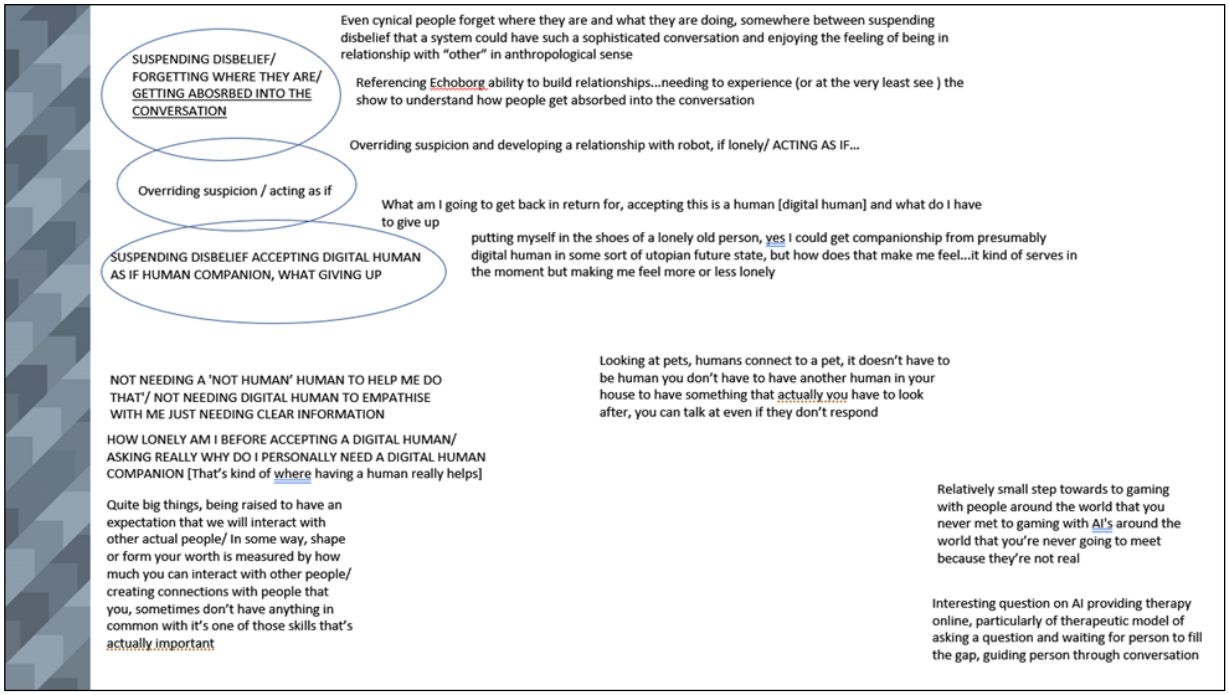
********Prompt:**

- **What is important about my research is… the unknown about AI social robots/avatars by the general public, yet how widespread and impactful they are likely to be in human society within the next 5-10 years. Research has indicated how the impact of t**
- **he internet and social media specifically, has affected humans' way of communicating, language and relating and relationships with others**

- Longer cultural trends way after the start of social media and now we seem to have to fumble at managing this, but as it seems to have become such an intrinsic part of our lives. It seems essential/ paramount to ask the question of AI social robots/avatars and how humans' experience them in interactional contexts, to work to develop a theory to support future potential outcomes of HRI (AIHI)…

- "There was a child that went forth every day. And the first object he looked upon, that object he became" (Walt Whitman)

(10/7/2022)



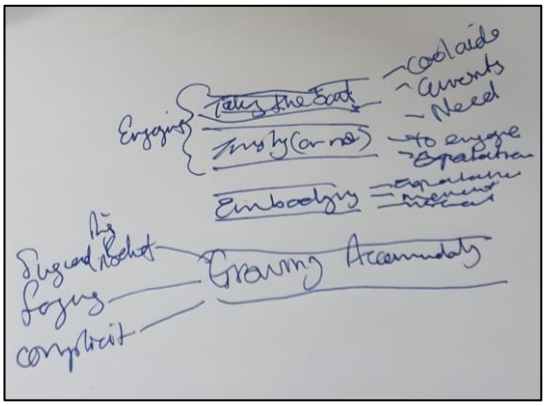
Appendix O: Memo on Theoretical Sampling and Categories (25/1/2022)

- In determining what the categories are, I need to determine the properties of the categories and their process and action.
- When looking at the human experience interacting with AI digital avatar (DA) and/or social robot (SR), I have become aware that there are Conditions and Contexts that impact/influence the expectation and possibly the experience or making sense of the experience
- (Re: Spinoza – Emotion in body comes first and then feeling in the mind). This seems important and speaks to what might be happening when people are 'being in the moment' with SR or DA.
- Reality how people feel?
- The theoretical sampling o[redacted]really explained what I had determined from previous interviews in parts, and then theoretical sampling of interview[redacted]as the creator of technolog[redacted]experienced, helped in the understanding of how they created that experience and what was intentional and what was not.
- (Theoretical) Extant literature (podcasts, James May, and Sherry Turkle) support the categories that have come out of the interviews. Where there has been a negative case, such Participant I, it raises the question of whether it is due to the personality of the participant or the quality of the technology that they didn't find the digital avatar helpful/engaging preferring people or a chatbot, or both?
- The social robots as a physically present technology (whether more human looking or machine looking seemed to receive more openness to being interacted with, and human looking one creating more uncanny valley and higher expectations.
- Why do some participants refer to SR or DA as their friend and others not?
- "Humans act towards things based on the meaning they have for them"
- How do they determine meaning? Through feeling? Projection? Need?
- Evolution psychology? Evolution of human social interactions? Evolution of attachment? Transitional objects?

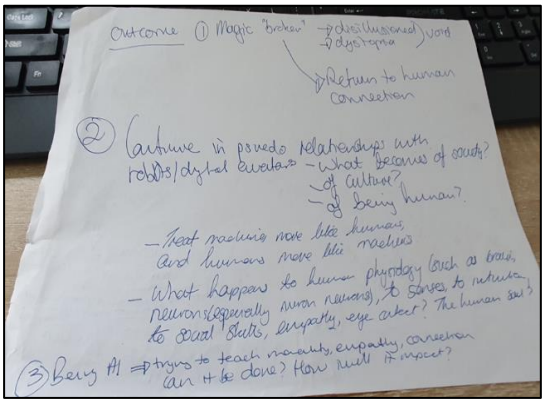
Partcipant_14 theoretical sampling

Appendix P: Examples of Progression of Diagrams

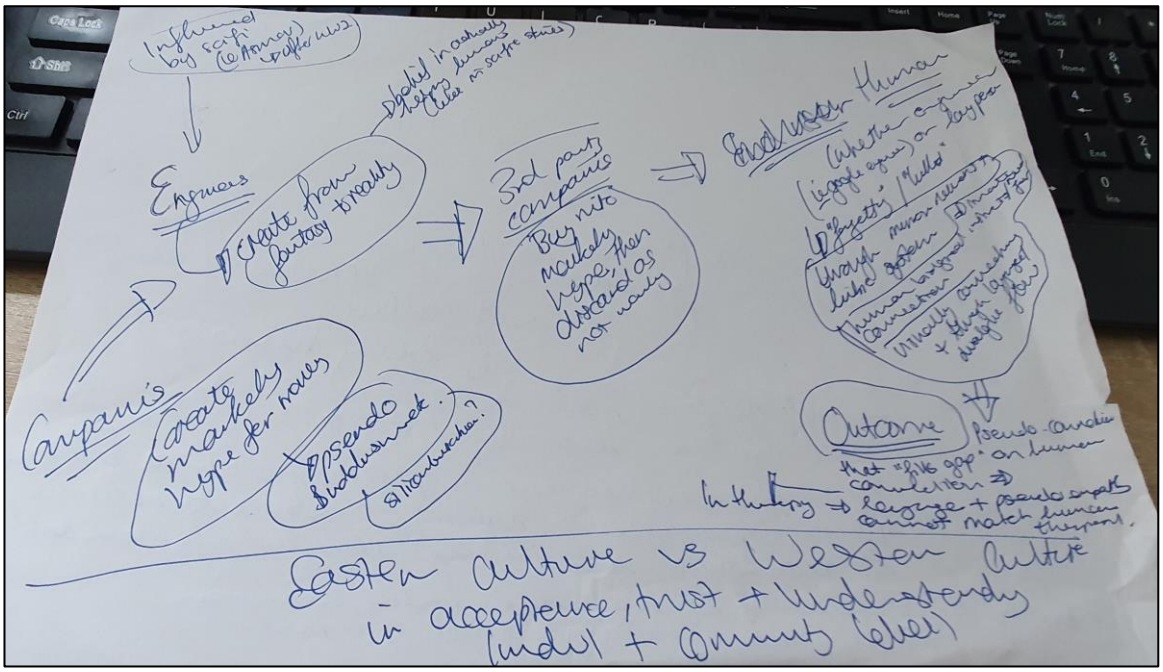
(4/5/2022)



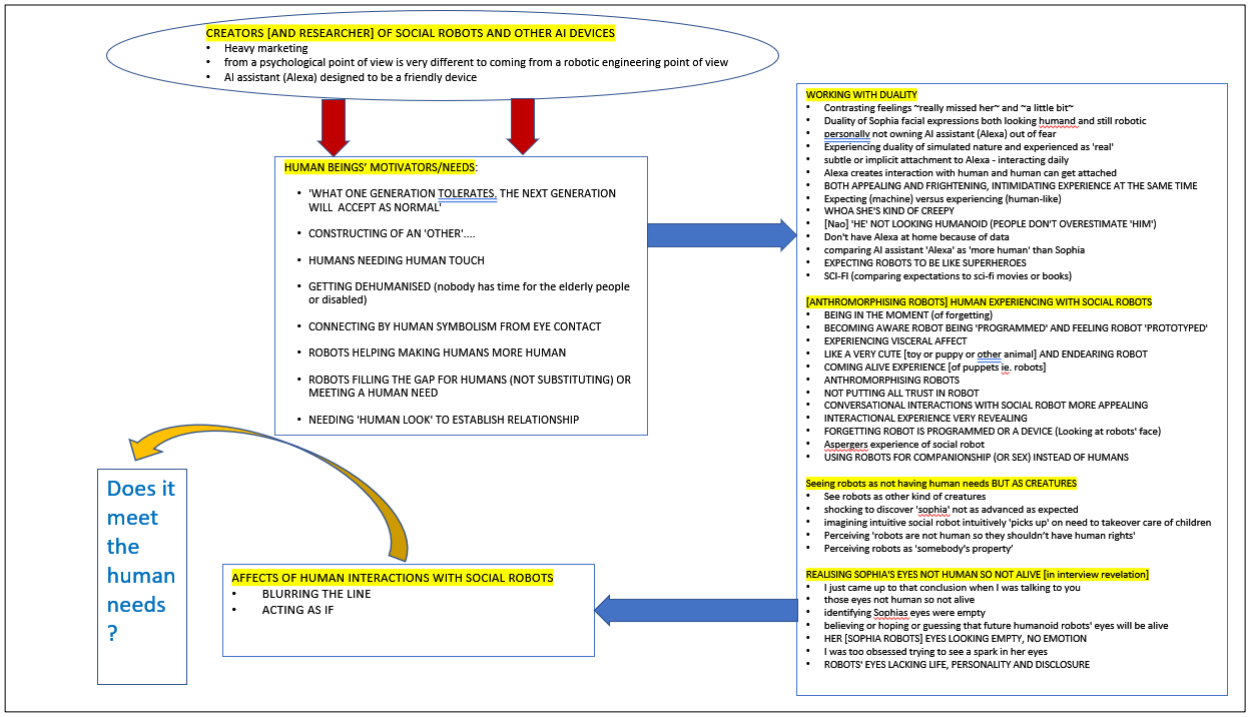
(28/5/2022)



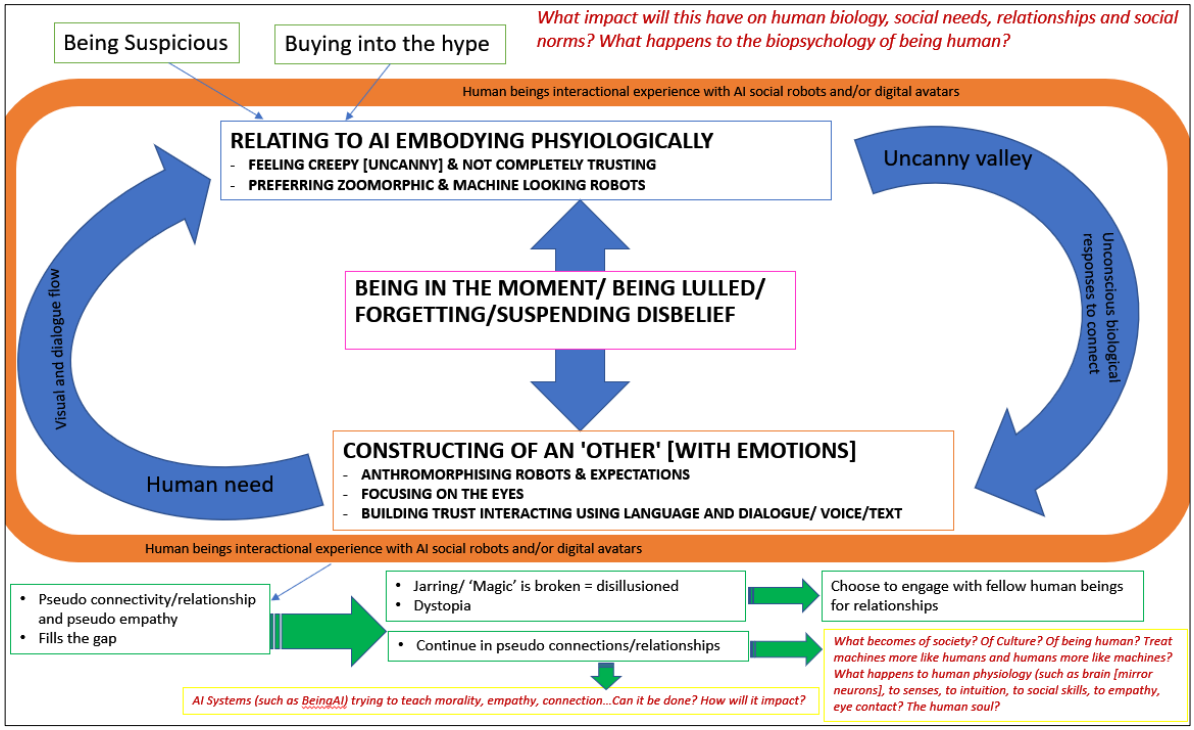
(21/6/22)



(28/2/22)


CREATORS [AND RESEARCHER] OF SOCIAL ROBOTS AND OTHER AI DEVICES
Heavy marketing
from a psychological point of view is very different to coming from a robotic engineering point of view
AI assistant (Alexa) designed to be a friendly device
HUMAN BEINGS' MOTIVATORS/NEEDS:
'WHAT ONE GENERATION TOLERATES, THE NEXT GENERATION WILL ACCEPT AS NORMAL'
CONSTRUCTING OF AN 'OTHER'....
HUMANS NEEDING HUMAN TOUCH
GETTING DEHUMANISED (nobody has time for the elderly people or disabled)
CONNECTING BY HUMAN SYMBOLISM FROM EYE CONTACT
ROBOTS HELPING MAKING HUMANS MORE HUMAN
ROBOTS FILLING THE GAP FOR HUMANS (NOT SUBSTITUTING) OR MEETING A HUMAN NEED
NEEDING 'HUMAN LOOK' TO ESTABLISH RELATIONSHIP
Does it meet the human needs ?
AFFECTS OF HUMAN INTERACTIONS WITH SOCIAL ROBOTS
BLURRING THE LINE
ACTING AS IF
WORKING WITH DUALITY
Contrasting feelings ~really missed her~ and ~a little bit~
Duality of Sophia facial expressions both looking humand and still robotic
personallv not owning AI assistant (Alexa) out of fear
Experiencing duality of simulated nature and experienced as 'real'
subtle or implicit attachment to Alexa - interacting daily
Alexa creates interaction with human and human can get attached
BOTH APPEALING AND FRIGHTENING, INTIMIDATING EXPERIENCE AT THE SAME TIME
Expecting (machine) versus experiencing (human-like)
WHOA SHE'S KIND OF CREEPY
[Nao] 'HE' NOT LOOKING HUMANOID (PEOPLE DON'T OVERESTIMATE 'HIM')
Don't have Alexa at home because of data
comparing AI assistant 'Alexa' as 'more human' than Sophia
EXPECTING ROBOTS TO BE LIKE SUPERHEROES
SCI-FI (comparing expectations to sci-fi movies or books)
[ANTHROMORPHISING ROBOTS] HUMAN EXPERIENCING WITH SOCIAL ROBOTS
BEING IN THE MOMENT (of forgetting)
BECOMING AWARE ROBOT BEING 'PROGRAMMED' AND FEELING ROBOT 'PROTOTYPED'
EXPERIENCING VISCERAL AFFECT
LIKE A VERY CUTE [toy or puppy or other animal] AND ENDEARING ROBOT
COMING ALIVE EXPERIENCE [of puppets ie. robots]
ANTHROMORPHISING ROBOTS
NOT PUTTING ALL TRUST IN ROBOT
CONVERSATIONAL INTERACTIONS WITH SOCIAL ROBOT MORE APPEALING
INTERACTIONAL EXPERIENCE VERY REVEALING
FORGETTING ROBOT IS PROGRAMMED OR A DEVICE (Looking at robots' face)
Aspergers experience of social robot
USING ROBOTS FOR COMPANIONSHIP (OR SEX) INSTEAD OF HUMANS
Seeing robots as not having human needs BUT AS CREATURES
See robots as other kind of creatures
shocking to discover 'sophia' not as advanced as expected
imagining intuitive social robot intuitively 'picks up' on need to takeover care of children
Perceiving 'robots are not human so they shouldn't have human rights'
Perceiving robots as 'somebody's property'
REALISING SOPHIA'S EYES NOT HUMAN SO NOT ALIVE [in interview revelation]
I just came up to that conclusion when I was talking to you
those eyes not human so not alive
identifying Sophias eyes were empty
believing or hoping or guessing that future humanoid robots' eyes will be alive
HER [SOPHIA ROBOTS] EYES LOOKING EMPTY, NO EMOTION
I was too obsessed trying to see a spark in her eyes
ROBOTS' EYES LACKING LIFE, PERSONALITY AND DISCLOSURE

(23/7/22)


Being Suspicious
Buying into the hype
What impact will this have on human biology, social needs, relationships and social norms? What happens to the biopsychology of being human?
Human beings interactional experience with AI social robots and/or digital avatars
RELATING TO AI EMBODYING PHSYIOLOGICALLY
- FEELING CREEPY [UNCANNY] & NOT COMPLETELY TRUSTING
- PREFERRING ZOOMORPHIC & MACHINE LOOKING ROBOTS
Uncanny valley
Unconscious biological responses to connect
BEING IN THE MOMENT/ BEING LULLED/ FORGETTING/SUSPENDING DISBELIEF
Visual and dialogue flow
Human need
CONSTRUCTING OF AN 'OTHER' [WITH EMOTIONS]
- ANTHROMORPHISING ROBOTS & EXPECTATIONS
- FOCUSING ON THE EYES
- BUILDING TRUST INTERACTING USING LANGUAGE AND DIALOGUE/ VOICE/TEXT
Human beings interactional experience with AI social robots and/or digital avatars
Pseudo connectivity/relationship and pseudo empathy
Fills the gap
Jarring/ 'Magic' is broken = disillusioned
Dystopia
Choose to engage with fellow human beings for relationships
Continue in pseudo connections/relationships
What becomes of society? Of Culture? Of being human? Treat machines more like humans and humans more like machines? What happens to human physiology (such as brain [mirror neurons], to senses, to intuition, to social skills, to empathy, eye contact? The human soul?
AI Systems (such as BeingAI) trying to teach morality, empathy, connection...Can it be done? How will it impact?