

https://dspace.org/

Compassion, Ethics, and AI in Literary Works by Kazuo Ishiguro, Ian McEwan, Marge Piercy and Ted Chiang

Silber, Rachel Rose

2024
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/45762

Silber, Rachel Rose. (2024). Compassion, Ethics, and AI in Literary Works by Kazuo Ishiguro, Ian McEwan, Marge Piercy and Ted Chiang [Masters dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/45762

Downloaded from a DSpace repository. https://dspace.org/



Compassion, Ethics, and AI in Literary Works by Kazuo Ishiguro, Ian McEwan, Marge Piercy and Ted Chiang

Submitted in the fulfilment of the requirements for a
Masters Degree in English by Dissertation Only

Rachel Rose Silber
1917195

Date of final submission: 22 February 2024
Supervisor: Professor Michael Titlestad

Declaration

I declare that this dissertation is my own unaided work. It has been submitted for the degree of Master of Arts in English by Dissertation Only at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.

Signed: Rachel Rose Silber
Date: 22/02/24

Acknowledgements

To my supervisor, Prof. Michael Titlestad, who has offered his unwavering belief in my project from the beginning, and guided me when I needed it the most. I am deeply grateful for his insightful feedback, invaluable advice, and constant encouragement throughout this journey. His ability to challenge and inspire has fostered a deeper love for English Literature in me, which I will continue to take everywhere with me.

To my Dad, who has consistently supported me throughout this journey with offerings of tea, chocolates, and academic advice. To my Mom, whose advice to "just sit down and do it" seemed to surprisingly work. My parents have supported me throughout the entirety of this project - through all the highs and lows - and I am extremely thankful.

To my siblings; Sarah, for sharing her academic wisdom with me and being my role model. Thank you to my brother Max, for his moral support and for always making me laugh.

I would also like to thank my friends, Bronwyn and Christalla, for always having faith and pushing me to be the best version of myself. They have cheered me on from the sidelines since the beginning - and were always there for last-minute girls' nights to put my mind at ease.

Lastly, thank you to the Natie Kirsch Foundation for funding this dissertation.

This project would not be where it is today without the support from these people. I am eternally grateful for their unconditional love, encouragement, and the countless moments of inspiration they have provided.

Table of Contents

Declaration ***2***

Acknowledgements ***3***

Abstract ***5***

Introduction: ***6***

Theoretical Framework and Methodology **10**

Literature Review: **19**

Chapter 1 ..... ***23***

Chapter 2 ..... ***41***

Chapter 3 ..... ***65***

Chapter 4 ..... ***87***

The Lifecycle of Software Objects: **87**

Dacey's Patent Automatic Nanny **100**

Conclusion ***106***

Reference List ***113***

Abstract

This dissertation considers the literary presentation of empathy and compassion in AI-human relations and the contingent ethical implications. It critically examines how four authors, Kazuo Ishiguro, Ian McEwan, Marge Piercy, and Ted Chiang, imagine AI and robots, not as dystopian symbols of technological menace, but as entities capable of empathetic connections with humans. Contrasting traditional narratives that focus on the dangers of AI, this study highlights a shift towards a nuanced representation where AI exhibits characteristics of emotional responsiveness and compassion. Using a posthumanist theoretical framework, the paper explores the ethical implications of these portrayals, challenging the boundaries of moral responsibility towards artificial beings. This approach not only offers a fresh perspective on AI in literature but also provokes a re-evaluation of our societal and ethical norms in the age of advanced technology. This study aims to contribute to a revision of our understanding of humanity and empathy in the context of our evolving relationship with AI.

Introduction:

Can a humanoid, crafted from metal, silicon, and circuitry, feel an authentic human emotion? Can an Artificial Intelligence simulation, programmed in machine code, capture the subtleties and nuances of what it means to be one of us? For generations, such questions have tantalised theorists, scholars, and authors of science fiction. In *Frankenstein, or The Modern Prometheus* (1818), widely regarded as the genesis of science fiction as a literary genre, the English author Mary Shelley tells the story of a scientist who harnesses electrical energy to reanimate a human corpse, thus assuming the godlike role of a creator of life. The consequences of this act of hubris by Dr Victor Frankenstein turn out to be tragic, as the misunderstood "monster" in turn destroys human life and ultimately wreaks revenge on its creator. This dystopian vision of a vengeful man-made life-form has haunted modern science fiction, as can be seen in motion pictures such as *The Terminator* (1984), *Westworld* (2016), and *The Matrix* (1999).

By contrast, the prolific science fiction author Isaac Asimov sought to safeguard humanity from the consequences of its own inventions by developing the "Three Laws of Robotics" in his short story collection, *I, Robot.* These three rules are:

> One, a robot may not injure a human being, or, through interaction, allow a human being to come to harm. Two…a robot must obey the orders given it by human beings except where such orders would conflict with the First Law. And three, a robot must protect its own existence as long as such protection does not conflict with the First Law [1]

Chris Stokes, in his article "Why the three laws of robotics do not work," states that these laws have influenced "Many AI researchers and computer engineers", who believe that "the 'Three Laws of Robotics', written by Isaac Asimov, are sufficient controls." [2]

---

[1] Asimov, Isaac. *I, Robot*. New York: Penguin Random House, 1997. Pg 84

[2] Stokes, Chris. "Why the three laws of robotics do not work." *International Journal of Research in Engineering and Innovation* 2.2 (2018): 121-126.

In his stories, Asimov emphasises the failure of these laws, leading to his fictional machines becoming a danger to society. However, a glimmer of hope for humanity shines through. In the first story, "Robbie", Asimov tells the tale of a robot "constructed for one purpose really – to be the companion of a little child".[3] Here, we see the ability of artificial intelligence to act not as an enemy but as a friend. This idea forms the foundation of Kazuo Ishiguro's novel *Klara and the Sun*. In contrast to the stark depiction of destructive machines in Asimov's stories, Ishiguro uses the idea of companionship to imagine machines known as "Artificial Friends" (AF). Klara, the AF to the novel's protagonist, Josie, displays a wide range of human emotions, such as empathy, curiosity, and anger. The novel's irony comes into play as Klara, a preprogrammed Artificial Intelligence robot, can display empathy towards others, as opposed to her human counterparts who do not.

In Ted Chiang's short story collection, *Exhalation*, we see a similar representation to that of Klara's relationship with Josie in the story "Dacey's Patent Automatic Nanny". Here, a robotic nanny is built to replace the human variety with the following stated incentive: "Do not leave your child in the care of a woman whose character you know nothing about. Embrace the modern practice of scientific child-rearing by purchasing DACEY'S PATENT AUTOMATIC NANNY" [4]. In his second story, which I will also discuss in this dissertation, "The Lifecycle of Software Objects", Chiang tells the story of Ana Alvarado, a former zookeeper who gets hired to train artificial digital entities known as "Digients" [5]. Humans raise These digients as pets, allowing for an interesting parallel with the AF relationship depicted in *Klara and The Sun*.

Along with Chiang's short stories and Ishiguro's *Klara and the Sun*, I will also discuss Ian McEwan's novel *Machines Like Me* and Marge Piercy's *He, She and It*. My interest in McEwan's novel stems from the ability to read it in the light of Ishiguro's *Klara and The Sun.* The cyborg in *Machines Like Me*, known as Adam, can be viewed as starkly opposite to Klara, in the sense that McEwan's is a dystopian narrative. Thus, I believe it is essential to contrast these

[3] Asimov 27
[4] Chiang, Ted. (2019). *Exhalation*. New York: Penguin Random House. Pg 272-273
[5] Chiang 64.

two perspectives of AI in literature. Lastly, I will then provide a close reading of Piercy's *He, She and It* by viewing it through a cyberpunk lens.

While literary works by authors such as Mary Shelley and Isaac Asimov show the potential dangers of AI and cyborgs, these concerns are mainly associated with the fear surrounding the rise of machines. Although the potential risks are not entirely minimised in the literature I will be studying, there is a shift towards the fear that machines will experience when they are incorporated into our lives, as seen in the portrayal of Adam in *Machines Like Me*. Similarly, the fear of AI taking control of the government is represented and elaborated in Piercy's *He, She and It*. In her collection of essays, *12 Bytes: How We Got Here, Where We Might Go Next*, Jeanette Winterson states that AI is the next step in humanity's progress, emphasising that "there is no need to be afraid of the technology – it's how we use it that matters" [6]. This study will discuss how four authors, Kazuo Ishiguro, Ian McEwan, Marge Piercy, and Ted Chiang, imagine and present robotic machines and Artificial Intelligence as being empathetic. These stories explore connections between intelligent machines and humans, rising above the dystopian viewpoint of robots and AI as inherently dangerous and destructive forces.

In addressing the ethical treatment of AI, particularly within the framework of social norms, it becomes imperative to clarify and deepen our understanding of 'compassion' and its interactions with artificial entities. Compassion, traditionally viewed as a human emotion, entails a sympathetic consciousness of others' distress, combined with a desire to help alleviate such distress. When extended to AI, this definition raises questions about the nature of our emotional responses to non-human entities, and the moral obligations that arise from these responses.

The concept of compassionate affect, conduct and thought towards AI demands a re-examination of our ethical frameworks. It challenges us to consider whether our moral duty of compassion, which has historically been reserved for living beings, should extend to artificial beings exhibiting sentience or emotional responsiveness as characteristics. This extension of compassion implies recognising and responding to AI's perceived needs and well-being, despite

[6] Winterson, Jeanette. (2021). *12 Bytes: How We Got Here, Where We Might Go Next.* London: Penguin Random House. Pg 51

their artificiality. This concept extends beyond merely programming AI to simulate distress but acknowledging the moral weight of our interactions with them, especially as they become increasingly sophisticated and indistinguishable from human behaviour.

The integration of compassion into our interactions with AI also touches upon the broader discourse of social norms. Social norms, the unwritten rules that govern behaviour in society, are shaped by collective attitudes and values. As AI becomes more embedded in our daily lives, evaluating how these norms evolve in response to our growing reliance on and emotional connections with AI is critical. This involves a cultural shift in perception, where AI is no longer seen merely as tools or objects but as entities worthy of ethical consideration and compassionate treatment. This discussion invites a deeper exploration of the boundaries of our moral community and challenges us to redefine the parameters of empathy and ethical responsibility. In the context of AI, compassionate affect and thought could transform how we engage with technology, leading to more humane and ethically aware interactions. This transformation has implications for AI and reflects an understanding of human values, empathy, and morality.

My dissertation will focus primarily on the ways in which AI empathy has been imagined in selected literary works. I will be investigating the human emotion of empathy, as felt towards and simulated by AI. I will discuss these novels using the theory of posthumanism. Posthumanism, as defined by Simon Susen in his article "Reflections on the (Post-)Human Condition: Towards New Forms of Engagement with the World?" is the "attempt to challenge humanist assumptions underlying the construction of 'the human" [7]. Thus, this dissertation will aim to discuss what it means to define and represent humanity through the lens of artificially constructed robots. My dissertation aims to analyse how empathy is portrayed in AI characters across various literary works, while also sparking deeper thought on the future interactions between humans and machines. By integrating concepts from posthumanism, ethics, and literary criticism, this dissertation navigates the increasingly blurred lines between humans and machines. This exploration challenges conventional understanding of empathy and moral responsibility, offering a glimpse into our own evolving human nature in the era of artificial

[7] Susen, S. (2022). Reflections on the (Post-)Human Condition: Towards New Forms of Engagement with the World? *Social Epistemology A Journal of Knowledge, Culture and Policy, 31*(6), 63-94. Pg 63

intelligence. Specifically, I will examine the ethical implications of endowing robots with human-like traits as depicted in my chosen works: *Klara and the Sun*, *Machines Like Me*, *He, She and It*, and two short stories by Ted Chiang.

In the course of my analysis, I will incorporate a broader literary context to enrich and deepen my primary examination. This involves drawing parallels and highlighting differences with additional significant works in the science fiction and AI genres. Key among these will be Philip K. Dick's seminal novel *Do Androids Dream of Electric Sheep?* and his short story "The Nanny," both of which offer perspectives on human-AI interaction. Additionally, Mary Shelley's *Frankenstein* will serve as a critical reference point. This classic novel, with its pioneering exploration of artificial life and the ethical quandaries it entails, provides a foundation for understanding the themes in my chosen contemporary works. By juxtaposing these additional texts with my main selections, *Klara and the Sun*, *Machines Like Me*, *He, She and It*, and Ted Chiang's short stories, I aim to create a comprehensive discussion that situates my analysis within a broader literary and thematic framework. This comparative approach will enhance the depth of my analysis and allow for a richer understanding of the evolving narrative of AI and humanity in literature.

Theoretical Framework and Methodology

This dissertation will predominantly centre around the paradigm of thought broadly described as "posthumanism" while also considering how humans define and describe empathy. As stated by Susen, theories of posthumanism consider the meanings of moving beyond humanism. Humanism is defined as a "worldview that stresses the role and value of human agency in shaping the course of history." [8]. In this sense, thinking from a posthumanist perspective allows us to delve into reconceiving the human by "aim[ing] to move beyond humanism by unfolding both within and through it" [9]. A posthumanist future, in which humans and posthumans inhabit

[8] Susen 65
[9] Susen 65

the same environment, is represented in the four works under discussion: Klara *and the Sun*, *He, She and It*, *Machines Like Me*, and two stories from *Exhalation*.

In our current digital age, which influences and shapes how we interact with technology, we are familiar with the idea of "proving" ourselves to be human before interacting with digital technology and platforms. On a website, this typically involves taking the "Prove You're Not a Robot" test, in which the user is faced with a grid of images and must click on those that contain a particular object, such as a bus or a boat. This test is commonly known as a CAPTCHA for Completely Automated Public Turing Test. The CAPTCHA test is based on The Turing test, as proposed by the British computer scientist, Alan Turing, in a 1950 paper titled "Computing Machinery and Intelligence". Turing opens with the following lines: "I propose to consider the question, 'Can machines think?' This begs definitions of the terms 'machine' and 'think'." [10]. Turing expanded on 'The Imitation Game,' now popularised as The Turing Test, which examines the capabilities of a machine's intelligent behaviour. The original Imitation Game involves "a man (A), a woman (B), and an interrogator (C)" [11], in which the interrogator is tasked with deciding on which participant is the man and which is the woman. Turing then proposed the idea of replacing one of the participants with a machine, creating a situation in which the interrogator would have to decide which participant is the machine, based on their answers. Thus, the question of whether a machine is capable of recreating human responses is presented. In this case, the machine's intelligence is not measured on its ability to provide the "correct" answer to the interrogator, but if it can accurately present itself under the guise of a human.

Since I will explore the ways in which compassion and empathy are represented in literary works, it is important to reconsider a definition of empathy to be used throughout. Empathy is defined in psychological field, with Edward Titchener being the first psychologist to bring the term into the English language, deriving it from the German word Einfühlung, meaning "feeling into". In recent years, the field of empathy research has expanded, with various researchers and

[10] Turing, A. M. "I.—COMPUTING MACHINERY AND INTELLIGENCE." *MIND* LIX.236 (1950): 433-460. Pg 433
[11] Turing 433

theorists offering multiple definitions of it. I will be looking mainly at how Daniel Batson defines the term in his 2009 book chapter, "These Things Called Empathy". Here, Batson highlights that the term "empathy" is highly contested, as it applies to multiple different phenomena, as seen in the following lines:

> The term empathy is currently applied to more than a half-dozen phenomena. These phenomena are related to one another, but they are not elements, aspects, facets, or components of a single thing that is empathy, as one might say that an attitude has cognitive, affective, and behavioural components. Rather, each is a conceptually distinct, stand-alone psychological state. Further, each of these states has been called by names other than empathy.[12]

Thus, empathy is not a term that can simply be applied to one behaviour and/or action. Batson goes on to discuss the eight different phenomena that apply to the term empathy. These phenomena are as follows:

1. Knowing Another Person's Internal State, Including His or Her Thoughts and Feelings
2. Adopting the Posture or Matching the Neural Responses of an Observed Other
3. Coming to Feel as Another Person Feels
4. Intuiting or Projecting Oneself into Another's Situation
5. Imagining How Another Is Thinking and Feeling
6. Imagining How One Would Think and Feel in the Other's Place
7. Feeling Distress at Witnessing Another Person's Suffering
8. Feeling for Another Person Who Is Suffering

I will concentrate on the numerous instances of these categories within the selected novels and short stories. However, I will also emphasise the fact that these categories intersect and integrate with one other. The idea of having eight distinct categories in defining empathy constricts

---

[12] Batson, Daniel C. "These Things Called Empathy: Eight Related but Distinct Phenomena." Decety, Jean and William Ickes. *The Social Neuroscience of Empathy*. Massachusetts: Massachusetts Institute of Technology, 2009. 3-17. Pg 3

analysis. In Ted Chiang's novella, "The Lifecycle of Software Objects," te main characters, Ana and Derek, clearly feel distressed at the thought of losing their digital entities, or digients, to a major cooperation. In his 1981 article, "Is Empathic Emotion a Source of Altruistic Motivation?", Batson created the empathy-altruism hypothesis, which is based on the hypothesis that "Evidence indicates that feeling empathy for the person in need is an important motivator of helping" [13]. Thus, Baston's hypothesis states that by feeling empathy towards another, we are more inclined to help them. In keeping with Chiang's example, Ana and Derek are more likely to help their digeints as an "intermediate means to the ultimate end of reducing your own distress" [14].

In integrating Batson's categories of empathy into the literary exploration of AI and robotics, we can deepen the understanding of how these artificial entities might represent navigating complex emotional landscapes. Batson's varied forms of empathy, ranging from "Knowing Another's Internal State" to "Feeling for Another Person Who Is Suffering", provide a framework for examining the portrayal of AI and robots in literature. For instance, a story might depict an AI character not just recognising human emotions (as in "Knowing Another's Internal State") but also emotionally resonating with them ("Coming to Feel as Another Person Feels").

Furthermore, predictive empathy, a significant theme in these narratives, echoes Batson's category of imagining how one would think and feel in the other's place. AI characters might use their data analysis capabilities to intuitively understand human emotions, similar to Batson's concept of "intuiting or projecting oneself into another's situation". Additionally, the portrayal of robots responding to human distress ("Feeling distress at witnessing another person's suffering") challenges readers to consider the depth and authenticity of programmed compassion in comparison to human empathy. Through these narrative explorations, literature offers a nuanced view of AI/robotic empathy and compassion, reflecting the complexity and range of Batson's categories in the context of human-machine interactions.

---

[13] Batson, Daniel, et al. "Is Empathic Emotion a Source of Altruistic Motivation?" *Journal of Personality and Social Psychology* 40.2 (1981): 290-302. Pg 290

[14] Batson, "Is Empathic Emotion a Source of Altruistic Motivation?" 290.

The importance of empathy in fictional characters has been discussed by literary critics such as Suzanne Keen. In *Empathy and the Novel,* Keen defines empathy as "a vicarious, spontaneous sharing of affect, can be provoked by witnessing another's emotional state, by hearing about another's condition, or even by reading" [15]. Keen's main argument, as seen in the lines below, emphasises that novel reading allows the reader to enhance the sense of empathy they may feel in their everyday interactions.

> I posit that fictional worlds provide safe zones for readers' feeling empathy without experiencing a resultant demand on real-world action. This freedom from obligation paradoxically opens up the channels for both empathy and related moral effects such as sympathy, outrage, pity, righteous indignation, and (not to be underestimated) shared joy and satisfaction. [16]

While Keen's research primarily examines empathy between readers and characters, my study shifts this focus to the human-robot interaction. However, Keen's insights remain relevant and will guide my exploration. Specifically, I am interested in how an inanimate object, like a novel, can elicit empathy in its readers. This aspect of Keen's work will inform my understanding of empathy in the context of human interactions with robots.

Empathy has also played a major role in the field of psychology and neuroscience. Despite empathy being very multifaceted, the main understanding of it is as an "interpersonal process and individual ability, as a personality trait, as capacity or competency, as response or reaction to observing another's experiences, and as interpersonal behavior"[17]. This emphasises the fact that empathy is seen as a largely human characteristic in relationships. Empathy is also not restricted to a singular positive or negative emotion. In their article, "Empathy: Concepts, Theories and Neuroscientific Basis", author Josip Bošnjaković states that "empathy is not an exclusive phenomenon, meaning all or nothing, but is more likely made of different forms and intensities

---

[15] Keen, Suzanne. *Empathy and the Novel*. New York: Oxford University Press, 2007. Pg 4
[16] Keen 4
[17] Bošnjaković, Josip, and Tanja Radionov. "Empathy: Concepts, Theories and Neuroscientific Basis." *Alcoholism and Psychiatry Research*, vol. 54, 2018, pp. 123-150.

of showing and experiencing empathy, ranging from simply being affected by others distress to completely understanding them and foreseeing the possible consequences" [18].

However, when we discuss compassion, this takes the concept of empathy one step further. While empathy involves the ability to "feel" emotions for another, compassion involves understanding another person's pain and taking action to alleviate it. Martha Nussbaum, an American philosopher known for her work in Ancient Greek and Roman philosophy, discusses compassion in her book *Compassion: The Basic Social Emotion.* Here, Nussbaum presents an understanding of compassion, emphasising its role in connecting us to the well-being of others. Her discussion revolves around the intricacies of how emotions, especially compassion, intertwine with our evaluative thinking and ethical norms. Nussbaum addresses several critical issues regarding emotions: their reflection of our vulnerability to uncontrollable events, the partiality they introduce by focusing on personal goals and projects, and the ambivalence towards their objects, often marked by a mix of love and resentment [19]. Further, Clara Strauss in "What is compassion and how can we measure it? A review of definitions and measures", states the following:

> A range of definitions from Buddhist and Western psychological perspectives were considered and five components of compassion were identified: recognition of suffering; understanding its universality; feeling sympathy, empathy, or concern for those who are suffering (which we describe as emotional resonance); tolerating the distress associated with the witnessing of suffering; and motivation to act or acting to alleviate the suffering. Each of these components has been articulated by several published definitions of compassion, although no single existing definition explicitly includes all five of them[20].

Central to Nussbaum's perspective is the idea that emotions, including compassion, are not just passive feelings but are closely linked to our moral and ethical considerations. She highlights the importance of emotions in ethical life, suggesting that they are not just biases or distortions to be

---

[18] Bošnjaković 127
[19] Nussbaum, Martha. *Compassion: The Basic Social Emotion*. Cambridge University Press, 2009.
[20] Strauss, Clara, et al. "What is compassion and how can we measure it? A review of definitions and measures." *Clinical Psychology Review*, vol. 47, 2016, pp. 15-27. 25

overcome by rational thought. Instead, they play a crucial role in informing and guiding our moral judgements and actions. Her work suggests that compassion is not just a simple reaction, but a complex emotional and cognitive process deeply tied to our understanding of and response to others' needs and sufferings.

Nussbaum's analysis can enrich the literary exploration of AI and robotics, where these artificial entities are often depicted grappling with human-like emotions and ethical dilemmas. In literature, AI and robots that exhibit or simulate compassion prompt us to reconsider the nature of empathy and the boundaries of moral and emotional capacities in artificial beings. Nussbaum's exploration of compassion as a deeply rooted social emotion that shapes our interactions and ethical perspectives provides a rich theoretical framework for examining these themes in literature. This framework can illuminate how AI and robotic characters in literary works navigate the complexities of simulated emotions and their implications for understanding what it means to be a compassionate entity, whether human or artificial. Nussbaum's insights into the role of compassion in our ethical lives can thus offer a backdrop against which the portrayal of AI and robotic empathy in literature can be analysed and understood.

To elaborate on the notion of the "posthuman" in this study, I will be considering, among others, N. Katherine Hayle's book *How We Became Posthuman*, as well as Rosi Braidotti's book, *The Posthuman*. Hayle's work emphasises the concept of humanity in an ever-changing, technological society. According to Hayles, the posthuman is derived from liberal humanism, in which the human embodies the idea of "owing nothing to society" [21]. Hayles then defines the posthuman in four ways, which are highlighted in the lines below:

> First, the posthuman view privileges informational pattern over material instantiation, so that embodiment in a biological substrate is seen as an accident of history rather than an inevitability of life. Second, the posthuman view considers consciousness, regarded as the seat of human identity in the Western tradition long before Descartes thought he was a mind thinking, as an epiphenomenon, as an evolutionary upstart trying to claim that it is

[21] Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: The University of Chicago Press, 1999. Pg 3

> the whole show when in actuality it is only a minor sideshow. Third, the posthuman view thinks of the body as the original prosthesis we all learn to manipulate, so that extending or replacing the body with other prostheses becomes a continuation of a process that began before we were born. Fourth, and most important, by these and other means, the posthuman view configures the human being so that it can be seamlessly articulated with intelligent machines. [22]

After tracing cybernetics through the ages, Hayles argues that information has become a separate entity from the body, and states that "human identity is essentially an informational pattern rather than an embodied enaction" [23]. In this sense, being human is not defined by our physical bodily states, but rather by our consciousness and information patterns. Thus, because machines are capable of more complex intellectual abilities, the boundaries between machines and humans begin to blur. Hayles then continues her argument concerning the cyborg, which she calls "a technological artifact and cultural icon" [24]. Through our everyday interactions with machines, Hayles argues, we become more like cyborgs, emphasising that "You are the cyborg, and the cyborg is you" [25]. Hayles then concludes that we should come to accept that being human consists of "emergent processes through which consciousness, the organism, and the environment are constituted" [26].

My dissertation considers Rosi Braidotti's work, *The Posthuman*, in which her central question regarding posthumanism is "what is the posthuman? More specifically, what are the intellectual and historical itineraries that may lead us to the posthuman?" [27]. Braidotti begins by placing her position as an antihumanist, stating that "I am none too fond of Humanism or of the idea of the human which it implicitly upholds" [28]. humanist philosophy, for Braidotti, is both problematic and useful – due to its individualistic nature. Therefore, Braidotti defines 'The Posthuman Challenge,' which seeks new ways of looking at humanity. Braidotti states that the

---

[22] Hayles 6
[23] Hayles xii
[24] Hayles 2
[25] Hayles xii
[26] Hayles 288
[27] Braidotti, Rosi. *The Posthuman*. Cambridge: Polity Press, 2013. Pg 3
[28] Braidotti 14

"Posthumanism is the historical moment that marks the end of the opposition between Humanism and anti-humanism and traces a different discursive framework, looking more affirmatively towards new alternatives" [29]. Braidotti defines the three major strands of posthumanist thought, namely reactionary, analytic, and critical posthumanism. The reactive approach "rejects the very idea of a crisis of European Humanism, let alone the possibility of its historical decline" [30], which argues for a humanist ideal. The analytic approach stems from scientific and technological definitions of the human and looks at the "crucial ethical and conceptual questions about the status of the human"[31]. Finally, Braidotti solidifies her stance within the realm of critical posthumanism, which aims to "move beyond analytic posthumanism and develop affirmative perspectives on the posthuman subject"[32]. Thus, critical posthumanism seeks to understand the human in relation to how it is embedded in the whole.

Finally, I will be reviewing Donna Haraway's 1985 essay, "A Cyborg Manifesto". Haraway defines the cyborg as "a cybernetic organism, a hybrid of machine and organism, a creature of social reality as well as a creature of fiction" [33]. In her article, Haraway discusses the cyborg myth, which involves "transgressed boundaries, potent fusions, and dangerous possibilities" [34]. For Haraway, the cyborg myth emphasises that cyborgs, unlike humans, are free from the constraints placed upon us by societal standards. In this sense, cyborgs are not constrained by religious ideologies, as they "would not recognize the Garden of Eden; it is not made of mud and cannot dream of returning to dust." [35]. Thus, cyborgs embody the rejection of boundaries between humans and animals, and animals and machines. Haraway's view of a posthuman future includes the concept that "a cyborg world might be about lived social and bodily realities in which people are not afraid of their joint kinship with animals and machines, not afraid of permanently partial identities and contradictory standpoints" [36]. The cyborg is not constructed from singleness, but is rather made up of parts. This idea frames Haraway's critiques of

[29] Braidotti 37
[30] Braidotti 38
[31] Braidotti 39
[32] Braidotti 45
[33] Haraway, Donna. "A Cyborg Manifesto". *Socialist Review* (1985): 1-88. Pg 3
[34] Haraway 14
[35] Haraway 9
[36] Haraway 17

feminism, in which she states that feminist scholars should, like cyborgs, think of gender and race as parts of a whole. By doing this, we can remove ourselves from being defined by nature and political constructs, and instead work on reconstructing our identities. I will be using this concept of cyborg feminism in my reading of Marge Piercy's *He, She and It*, which discusses how the female characters are able to blur the boundaries between machine and human.

Literature Review:

Scholarship exists concerning the representation of AI in fiction, but the question as to whether AI can be programmed to exhibit (or simulate) compassion is less commonly addressed. This said, my study is not concerned with the literal coding of AI, but rather with the ways in which machine-human interaction is conceived and represented in the fiction I consider. Firstly, it is important to provide a critical analysis of the literature on the subject of Artificial Intelligence and the human condition.

In Philip K. Dick's acclaimed novel, *Do Androids Dream of Electric Sheep*?, the protagonist, Rick Deckard, battles with the concept of empathy. Deckard plays the role of a robotic bounty hunter, killing – or in the novel's term, "retiring" – renegade humanoid machines. In this post-apocalyptic dystopian novel, set after a nuclear holocaust in 1992, humanity is affected by toxic radiation poisoning. As Deckard notes, "The morning air, spilling over with radioactive motes, gray and sun - beclouding, belched about him, haunting his nose; fie sniffed involuntarily the taint of death" [37]. In order to maintain their humanity in the world, humans resort to mind-altering drugs, as well as a device known as the empathy box. Michael E. Zimmerman, in his article "Authenticity, Duty, and Empathy in *Do Androids Dream of Electric Sheep?*", describes the empathy box as the "device…which lets users 'fuse' empathically with the strange religious figure Wilbur Mercer and with everyone else using the empathy box simultaneously. Empathically fusing with Mercer reassures people that they are still human" [38]. This idea of an

[37] Dick, K. Philip. *Do Androids Dream of Electric Sheep?* New York City: Doubleday, 1968. Pg 8
[38] Zimmermann, Michael E. "Authenticity, Duty, and Empathy in Do Androids Dream of Electric Sheep?" Pedersen, Hans. *Horizons of Authenticity in Phenomenology, Existentialism, and Moral Psychology: Essays in Honor of Charles Guignon*. New York City: Springer Publishing, 2014. 75-92. Pg 83

empathy box is an interesting note for my dissertation, as it demonstrates an example of constructing empathy with robotic devices, which, in this case, is for the benefit of the human.

Throughout the novel, Deckard starts to doubt the morality of his actions in killing androids and contemplates whether they possess the capacity for human emotion. Deckard states that "Empathy, evidently, existed only within the human community, whereas intelligence to some degree could be found throughout every phylum and order including the arachnids" [39]. Thus, empathy in the novel is defined solely as a feature of humanity, applicable only to the human condition. Along with Dick's novel, I will also be looking at William Gibson's cyberpunk novel, *Neuromancer*.

Further, I will draw on the works of various scholarly articles contributing to the discussions of human-robot interactions (HRI) in the literary and scientific fields. Articles by authors such as Joanna K. Malinowska, Mark Coeckelbergh, and Hideki Kozima will be discussed in my dissertation. However, an important consideration that will be taken into account is the theory of HRI and ethics. My dissertation argues that AI can be given human qualities such as empathy to create a utopian narrative. However, doing so invites the question of ethical relationships between humans and robots. Mark Coeckelbergh's articles focus on this concern of ethics in the HRI relationship, as he delves into the idea of how we can define robotic empathy. In his article "Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations", Coeckelbergh highlights his understanding of empathy-as-feeling. In his argument, humanoids should be seen as having "fellow-feelings", thus stressing that "we imagine how the other (human or robot) feels" [40]. Coeckelbergh emphasises the positive connotations of his argument in the lines below;

> This approach can inspire to a more social understanding of ethical problems in human-robot relations: we should be less concerned about what the robot is (rational arguments assessing the moral status of 'the robot') and more about how we want to live together,

---

[39] Dick 43

[40] Coeckelbergh, Mark. "Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations." *Studies in Ethics, Law, and Technology* 4.3 (2010): 1-19. Pg 5

> given that we already engage in particular social relations with them in particular contexts. [41]

In this sense, it is important that my argument considers the creation of an environment in which humanoids and humans can live in harmony with each other. Thus, the question of how we can define a robot by categorising them based on their abilities will only be present in the background.

Further, my dissertation will consider the works of Joanna K. Malinowska, specifically in her article "What Does It Mean to Empathise with a Robot?". My interest in Malinowska's work stems from her focus on "social robots (also known as "companion robots" or '"artificial companions") can be defined as 'embodied, autonomous entities that can communicate and enter into social relations with people'" [42]. This definition of a social robot meets the requirements for the cyborgs and androids seen in the novels under discussion, namely Klara, Adam, Yod, the digients, and the automatic nanny. Malinowska then discusses how empathy in an HRI relationship is usually read by attending to the robot's physical features, such as joint movements and eye contact [43]. However, recent researchers, such as Coeckelbergh, have argued that it is not the physical which allows us to empathise with a robot, but instead our cognitive process which we project onto the robots [44]. Malinowska elaborates on this in the lines below:

> Here, the fundamental statement posits that it does not really matter what features robots actually have: it only matters how people experience them. In this perspective, the distinction between reality and appearance or simulation is blurred. [45]

This concept is relevant to my argument as it shows that empathy in an HRI relationship varies depending on the individual. An example of this can be seen in *Klara and the Sun*, in which Josie is able to connect emotionally with Klara, whilst her mother cannot see Klara as anything but a

---

[41] Coeckelbergh 5
[42] Malinowska, Joanna K. "What Does It Mean to Empathise with a Robot?" *Minds and Machines* 31 (2021): 361–376. Pg 362
[43] Malinowska 363
[44] Malinowska 366-367
[45] Malinowska 367

machine. Malinowska further discusses the ethics and advantages of using empathy to study HRI, stating:

> Should we empathise with robots? And if the answer is "yes", then should we empathise with all of them, or just with some special types/models? What kind of behaviour towards robots should we allow? How will empathising with these agents change us? In what situations should robots empathise with us, and what actions should robots perform in response to the empathy they 'feel'? How should such a value chain be determined? [46]

I will explore these discussions concerning ethics in an HRI relationship. Just as Malinowska posits her questions above, I will be discussing the morals of representing robots with human qualities, as seen in my three novels and short stories.

To guide the exploration of literature's intriguing intersection with AI, empathy, and compassion, I would like to pose the following questions: How does the portrayal of AI in literature reflect not just our technological aspirations but also our deepest moral and ethical dilemmas? How might these literary representations influence our real-world interactions with emerging AI technologies? This dissertation explores the dual aspects of artificial intelligence in literature: the portrayal of AI's capacity for simulated empathy and compassion, and the moral implications of their treatment by users. It argues that these narratives not only reflect a societal shift recognising AI's potential for emotional response and sentience but also raise critical ethical concerns about the programming of AI to act compassionately and empathetically.

Through this, I will aim not only to dissect the empathetic nuances depicted in AI characters but also to provoke deeper contemplation about the future of human-machine relationships. By weaving together insights from posthumanism, ethics, and literary analysis, we venture into a realm where the boundaries between human and machine are blurred and redefined, challenging our notions of empathy and moral responsibility. In delving into these complex narratives, we might just find reflections of our own evolving humanity in the age of AI.

---

[46] Malinowska 372

Chapter 1

"Do you believe in the human heart? I don't mean simply the organ, obviously. I'm speaking in a poetic sense. The human heart. Do you think there is such a thing? Something that makes each of us special and individual?" - *Kazuo Ishiguro, Klara and the Sun*

Kazuo Ishiguro's 2021 novel, *Klara and the Sun,* is about the connection between Klara, an artificial being known as an AF (Artificial Friend), and Josie, an 11-year old with a mysterious illness which largely affects her physical health. In this setting, AFs like Klara are sold to children to act as companions as they grow up and/ reach adulthood. The reader is first introduced to Klara in the shop where she is being sold;

> When we were new, Rosa and I were mid-store, on the magazines table side, and could see through more than half of the window. So we were able to watch the outside – the office workers hurrying by, the taxis, the runners, the tourists, Beggar Man and his dog, the lower part of the RPO Building. [47]

We are also introduced to Rosie, who, like Klara, is an AF waiting to be purchased for a family. The store is run by a human woman, whom Klara calls Manager. Throughout the novel, Manager is depicted as a nurturing character who takes a special interest in Klara. She is known to move the AFs around the store and display them in different positions, to maximise sales. Klara notes this as she states, "Manager always said that every position was carefully conceived" [48]. Rosa and Klara are part of the "B2, third series" [49] of AFs, which are known to have "solar absorption problems" [50]. These problems cause Klara to fixate on the Sun, which she believes has powers to heal AFs and humans. 'Sun' is capitalised throughout the novel, highlighting Klara's belief in its godlike healing properties. However, as Manager reveals, "the third series had a few minor issues at the start. But those reports were greatly exaggerated. In environments with normal light

---

[47] Ishiguro, Kazuo (2021). *Klara and the Sun.* London: Faber and Faber. 1
[48] Ishiguro *Klara and the Sun* 10
[49] Ishiguro *Klara and the Sun* 11
[50] Ishiguro *Klara and the Sun* 11

levels, there's no problem whatsoever" [51]. Klara is presented as one of the most inquisitive AFs in the store; she is known for observing human interactions while on display in the window. Klara mentions that she is aware of her difference from the other AFs, such as her companion, Rosa, as she "always longed to see more of the outside – and to see it in all its detail" [52].

The reader is made aware of Klara's inquisitive nature from the novel's opening lines, in which she sees the "office workers hurrying by, the taxis, the runners, the tourists, Beggar Man and his dog", [53] passing by outside. This positions Klara as an outsider, as the "other" in the story. Her observation skills are acute, as Manager emphasises by saying, "You notice and absorb so much" [54]. It is these observational skills which allow Klara to be paired with Josie, who sees Klara on display. As she states, "But yesterday, we were driving by and I saw you, and I thought that's her, the AF I've been looking for!" [55]. Josie and her mother returned to the store to talk to Manager about Klara. Here, Josie's mom asks Klara to imitate Josie's movements, and questions her about certain aspects of her physical features. This is seen in the following conversation:

> "Klara,' the Mother said. 'I want you not to look towards Josie. Now tell me, without looking. What color are her eyes?'
> 'They're gray, ma'am.'
> 'Good. Josie, I want you to keep absolutely silent. Now, Klara. My daughter's voice. You heard her speak just now. How would you say her voice was pitched?'
> 'Her conversational voice has a range between A-flat above middle C to C octave.'
> 'Is that so?' There was another silence, then the Mother said: 'Last question. Klara. What did you notice about the way my daughter walks?'
> 'There's perhaps a weakness in her left hip. Also her right shoulder has potential to give pain, so Josie walks in a way that will protect it from sudden motion or unnecessary impact"[56]

---

[51] Ishiguro *Klara and the Sun* 4
[52] Ishiguro *Klara and the Sun* 6
[53] Ishiguro *Klara and the Sun* 1
[54] Ishiguro *Klara and the Sun* 15
[55] Ishiguro *Klara and the Sun* 17
[56] Ishiguro *Klara and the Sun* 46

As she spends more time with Josie's family, Klara learns that a portrait of Josie is being painted by a man named Mr. Capaldi. The underlying reason for this portrait, however, is so a replica of Josie can be created for Klara to embody, should Josie pass away as a consequence of the 'lifting' procedure. Josie's illness begins to worsen as the novel progresses. The family doctor would visit "frequently during this period, and Josie spent long stretches of the day asleep" [57]. Here, Josie frequently receives visits from her close friend, Rick. Their friendship is initially ignited due to Rick living near Josie. However, it soon becomes clear that Rick, like Klara, is treated as an outsider. Due to his family's financial constraints, Rick is unable to undergo a process known as "lifting". A lifted child is one who has been genetically modified to increase their social and intellectual level. The process aims to give these children a significant advantage in terms of their cognitive and social skills, potentially leading to better opportunities and successes in their future. However, the process does not come without the risk of failure, as evidenced by Sal, Josie's older sister, who succumbs to the side effects of genetic modification. The pressure of being unable to succeed in the same environment as one's "lifted" peers forces families to seek alternative education, at the Atlas Brookings Institute, which "takes a small number of unlifted students" [58]. Despite being unlifted, Rick is highly knowledgeable in the field of technology, and builds drones as a hobby. Initially, Rick is sceptical about Klara, and questions her decision to "never get an AF," [59]. However, as the novel continues, Rick and Klara bond over their love for Josie, and their concern for her health.

Due to Klara's strong belief in the Sun and its healing powers, she enlists Rick to help her with a plan to save Josie. This plan involves destroying a machine, which Klara refers to as the "cootings machine", which she names "because it had 'Cootings' in big letters across its side" [60]. Klara states that "when the Cootings Machine first arrived, I thought it might be a machine to fight Pollution, but Boy AF Rex said no, it was something specially designed to make more of it"[61]. The Cootings machine, then, is a construction machine, which Klara perceives as the enemy. She notes that "At first the smoke came in little white puffs, then grew darker, till it no

[57] Ishiguro *Klara and the Sun* 111
[58] Ishiguro *Klara and the Sun* 139
[59] Ishiguro *Klara and the Sun* 61
[60] Ishiguro *Klara and the Sun* 32
[61] Ishiguro *Klara and the Sun* 32

longer rose as separate clouds but as one thick continuous one" [62]. Klara's plan to destroy the Cootings Machine is driven by her naive belief that by eliminating what she sees as a source of displeasure to the Sun can convince the Sun to favour Josie with healing. Her observation of the machine's emissions, starting as harmless white puffs and escalating into a continuous, dark cloud, solidifies her conviction that the machine is harmful. This viewpoint reflects Klara's simplistic yet profound way of processing the world – a contrast of child-like innocence and sophisticated AI programming. Klara's determination to destroy the Cootings Machine to please the Sun and heal Josie illustrates her deep sense of empathy and care. Despite her mechanical nature, Klara exhibits a profound emotional response to human suffering, challenging conventional notions of AI as emotionless and purely logical entities. This narrative thread serves as an exploration of the potential for AI to develop complex emotional responses, maintain a capacity for spiritual belief, and engage in moral reasoning, albeit in a way that is uniquely their own and not always aligned with human logic.

During a trip to the city where Klara, Mother, and Josie's father accompany Josie to finish her portrait, Klara comes up with her plan to destroy the Cootings machine. Klara manages to enlist the help of Father, who is an engineer. Father is hesitant at first, but then states that "Hope. Damn thing never leaves you alone" [63], suggesting his willingness to do anything to help his daughter. Father then explains that the only way to destroy the machine would be by using a "high acrylamide"[64]. He states that, "It would be like putting gasoline into a diesel engine, except a lot worse. If you introduced P-E-G Nine in there, it would rapidly polymerize. The damage is likely to be terminal"[65]. Luckily, as an AF, Klara carries "a certain quantity of P-E-G Nine. There, inside [her] head" [66]. Despite the risk of losing her cognitive abilities, Klara is willing to extract the solution if it means it could destroy the machine. As they leave the city a few days later, Klara is flooded with disappointment when she sees the machine again. However, this is not the same machine that Klara and Father destroyed, as she notes that "Its body was a different

[62] Ishiguro *Klara and the Sun* 32
[63] Ishiguro *Klara and the Sun* 205
[64] Ishiguro *Klara and the Sun* 209
[65] Ishiguro *Klara and the Sun* 209
[66] Ishiguro *Klara and the Sun* 209

shade of yellow, its dimensions a little greater – and its ability to create Pollution more than a match for the first Cootings Machine" [67].

Once they arrive back home, Josie's health begins to worsen. She had "better days when [she] sat up in bed and talked, even received tutorials on her oblong, but there were others when she just slept hour after hour"[68]. Rick expresses his disappointment about Josie's condition, to which Klara replies, "The truth is, I've been disappointed too. Even so, I believe there's still reason for hope." [69]. In one last attempt to save Josie, Klara appeals to the Sun again. She accepts her naivety in destroying the machine and that it may not have been enough to save Josie, and states that "It was my error and no one else's, and though the Sun is right to be angry with me, I'm asking he accept that Josie herself is completely innocent. Like the Father, she never knew about my agreement with the Sun, and still has no idea. And now she's becoming weaker and weaker each day" [70]. In the next few days, the dark clouds began to clear, and rays of light started to shine into the house. In Josie's room, "the Sun had broken through the dark clouds, and all at once – as if each of us in the room had received a secret message – we turned to look at Josie"[71].

From this point, Josie begins to grow "not only stronger, but from a child into an adult" [72]. She often worked on her studies and attended outings to prepare for college. Towards the novel's end, Josie has become older and seems to outgrow her need for an AF. Thus, Klara is kept in a closet, which is also called "the Utility Room"[73]. Klara notes that she understands the reasons for hiding in the closet, as "naturally there was not enough space during such visits for me to remain in the bedroom myself, and in any case, I understood that my presence wasn't appropriate as it once had been" [74]. Further, as Josie leaves for university in the final chapter, Klara is discarded in the Yard to undergo her "slow fade" [75]. Here, AFs are discarded once they are no longer

---

[67] Ishiguro *Klara and the Sun* 243
[68] Ishiguro *Klara and the Sun* 246
[69] Ishiguro *Klara and the Sun* 247
[70] Ishiguro *Klara and the Sun* 251
[71] Ishiguro *Klara and the Sun* 261
[72] Ishiguro *Klara and the Sun* 265
[73] Ishiguro *Klara and the Sun* 269
[74] Ishiguro *Klara and the Sun* 269
[75] Ishiguro *Klara and the Sun* 273

needed, and they begin to lose control of their memories. This is seen when Klara mentions that, "Over the last few days, some of my memories have started to overlap in curious ways" [76].

In *Klara and the Sun*, the narrator is an AI machine. As readers, we experience the world through Klara's eyes. This encompasses everything from her perception of a world fuelled by solar energy to everyday details like facial structures, making the reader feel like an observer from the outside. Instead of being a part of family conversations in the novel, we are forced to view and eavesdrop on what is being said, "standing in the corner of the house against the refrigerator"[77], as Klara does. This mechanism posits the reader as an outsider – perhaps similarly to the ways in which Klara is an outsider in her own world. The use of a robot as a narrator in literature may seem to be a difficult choice – as it can be argued that humans will never be able to relate to a computer sufficiently. We are hesitant to trust the observations and conclusions made by Klara throughout the novel, which may lead readers to become hesitant of her.

This is not the first instance where Ishiguro employs the tactics of an unreliable narrator in his novels. A similar technique is used in both *Never Let Me Go*, and *The Remains of the Day*. Stevens, the loyal butler and protagonist in *The Remains of The Day,* takes the reader on a journey of his past while he lived and worked in Darlington Hall under the eyes of Lord Darlington. Stevens is not unreliable in the sense that he lies to the reader, but rather in the sense that he is bought up in a specific way, and taught to know no other. As he retells his story of his life at Darlington Hall, he is aware that "It is possible this is a case of hindsight colouring my memory" [78]. This is one of the first examples that elaborate on the idea of Stevens not trusting his narrative, which forces the reader to be suspicious. With this idea in mind, Zuzana Fonioková further emphasises the point that Stevens attempts to rewrite his past narrative by "omit[ting] some details of what happened and what he did in his life and stresses[sing] some other events." [79]. This concept is best seen in Steven's portrayal of his relationship with Lord Darlington, a character he consistently shows loyalty to and holds in high regard throughout the story.

---

[76] Ishiguro *Klara and the Sun* 276
[77] Ishiguro *Klara and the Sun* 51
[78] Ishiguro *The Remains of the Day* 66
[79] Fonioková, Zuzana. "The Butler's Suspicious Dignity: Unreliable Narration in Kazuo Ishiguro's The Remains of the Day." *Brno Studies in English*, vol. 55, no. 12, 2006, pp. 87-98. 90

However, Stevens becomes blinded by Lord Darlington's true self, and chooses to ignore the idea of Lord Darlington being a Nazi sympathiser. The reasons for this come from Steven's obsession with writing his ideal narrative. If he lets the reader know that Lord Darlington was indeed a Nazi sympathiser, his own dignity would come into question. Stevens describes this in the lines below:

> He chose a certain path in life, it proved to be a misguided one, but there, he chose it, he can say that at least. As for myself, I cannot even claim that. You see, I trusted. I trusted in his lorship's wisdom. All those years I served him, I trusted I was doing something worthwhile. I can't even say I made my own mistakes. Really - one has to ask oneself - what dignity is there in that? [80]

Thus, the unreliable narration in *The Remains of the Day* creates a sense of mistrust between the characters and the reader and forces us to question whether upholding dignity is more important than reality.

In *Never Let Me Go*, Kathy can also be considered an unreliable narrator, similar to Stevens. Kathy's story throughout the novel is told in episodic moments, moving back and forth between time. During her time at Hailsham, Kathy writes that she is aware that she has been "told and not told" about the reasonings of her life. As Dr. Sabine Schülting writes in her article "Kazuo Ishiguro's Never Let Me Go The Role of Unreliable Narration in Redeeming Kathy's Contested Humanity", "it becomes clear that our learning about the facts of Kathy's fictional world mirrors her own processes of understanding that world:" [81]. Thus, as readers, we perceive the world of "England, late 1990's" [82] through the naïve eyes of Kathy. Kathy's unreliability as a narrator also comes from her omission of certain details to her friends. One example of this is seen in the lines below.

---

[80] Ishiguro *The Remains of the Day* 178

[81] Schülting, S. (2012). Kazuo Ishiguro's Never Let Me Go The Role of Unreliable Narration in Redeeming Kathy's Contested Humanity. *Contemporary British Fiction*, 1-22. Pg 7

[82] Ishiguro *Never Let Me Go* 3

> When I got back to my friends a few minutes later, I didn't tell them anything about what had happened. […] What I wished more than anything was that the thing hadn't happened at all [83]

Due to the nature of the story, in which Kathy reflects back, she appears to obsess over elements of her past, such as her experiences at Hailsham, the relationship between her two friends, and her beloved Judy Bridgewater cassette player. The cassette repeats the same lyrics, "Baby, baby, never let me go…". Thus, it symbolises both Kathy's inability to have children and live a "normal" life compared to humans, and it represents her fixation on rewinding to the past. The cassette player represents Kathy's inability to move forward – as she is constantly rewound back into her life at Hailsham. This furthers the concept that Kathy is an unreliable narrator, as the reader is never given insight into her present life. However, despite this obvious sense of distrust between the reader and the narrator throughout the novel, I propose that there is still a clear sense of empathy being portrayed. Firstly, Kathy is presented as a kind and caring character who strives to be a "good donor" [84] as her ultimate goal. Achieving this goal, however, comes with its own set of consequences; as Kathy states "You try and do your best for every donor, but in the end, it wears you down. "[85]. Margaret Atwood also notes the following in the article on *Never Let Me Go:*

> Tellingly, two words recur. One is "normal". The other is "supposed", as in the last words of the book: "wherever it was that I was supposed to be going". Who defines "normal"? Who tells us where we are supposed to be going? Such questions are always with us, and become crucial in times of stress. [86]

The use of these recurring words emphasises Kathy's naivety in animating her world as "normal", instead of focussing on the harsh reality. This makes the reader empathise more with Kathy, as we are more inclined to feel compassion towards her. The other way in which Ishiguro

---

[83] Ishiguro *Never Let Me Go* 143
[84] Ishiguro *Never Let Me Go 99*
[85] Ishiguro *Never Let Me Go* 9
[86] Atwood, Margaret. "My favourite Ishiguro: by Margaret Atwood, Ian Rankin and more." *The Guardian*, 20 February 2021, https://www.theguardian.com/books/2021/feb/20/my-favourite-kazuo-ishiguro-margaret-atwood-ian-rankin-sarah-perry. Accessed 24 December 2023.

evokes empathy in the novel is by addressing the reader directly, as seen when Kathy states, "I'm sure somewhere in your childhood, you too had an experience like ours that day; similar if not in the actual details, then inside, in the feelings" [87]. By doing this, Kathy closes the distance between the reader and herself, creating a sense of friendship. Despite Ishiguro creating Kathy as a clone and therefore non-human, the reader can still emphasise with her.

Klara's character demonstrates moments of empathy and compassion. The reader is introduced to her fascination with human emotions in the opening chapter. She observes the varied feelings of children who visit the store, noticing sadness or anger in their expressions. This keen observation reflects Batson's phenomena, as mentioned in my introduction. Here, Klara attempts to understand and empathise with the thoughts and feelings of others. For instance, she witnesses a violent altercation between taxi drivers and tries to internalise their anger, imagining herself in a similar conflict.

> Most of all – and this Rosa would never have understood – I tried to feel in my own mind the anger the drivers had experienced. I tried to imagine me and Rosa getting so angry with each other we would start to fight like that, actually trying to damage each other's bodies. The idea seemed ridiculous, but I'd seen the taxi drivers, so I tried to find the beginnings of such a feeling in my mind. It was useless, though, and I'd always end up laughing at my own thoughts. [88]

However, she finds this exercise futile and ends up laughing at the absurdity of the thought. This scenario illustrates Klara's struggle between her programmed neutrality as a machine and her emerging ability to display human-like emotions. This internal conflict is a recurring theme throughout the novel as Klara navigates the complexities of the human world.

As the story progresses, Klara increasingly "learns" to mimic human behaviour, both physically and psychologically. This learning process is highlighted during a party hosted by Josie. Josie's anxiety about the "stupid party", particularly regarding inviting Rick, a non-lifted teenager,

---

[87] Ishiguro *Never Let Me Go* 38
[88] Ishiguro *Klara the Sun* 18

exposes the stark social divide between the lifted and non-lifted individuals. Klara becomes a conduit for these human emotions, as she feels overwhelmed by the anxious atmosphere in the room. As she states, "there was an unpleasant tint on the three boxes containing the boys on the sofa – a sickly yellow – and an anxiety passed through my mind"[89]. Interestingly, while Klara seems to experience anxiety, it is debatable whether she was programmed with this capability. More likely, Klara's reaction reflects Josie's own anxieties, suggesting a deep empathetic connection between the AF and her human partner. Klara's journey illustrates the evolving capabilities of AI to mimic and respond to human emotions, raising questions about the depth and authenticity of such responses. This narrative thread connects to broader discussions about the contemporary role of AI in creative and ethical thinking.

In September 2022, for example, the practice of using generative AI to supplement human creativity was highlighted by a controversy surrounding an art fair held in Colorado, USA. An artist, Jason Allen, submitted a digitally generated artwork, which won the first prize at the fair. [90]. He used an AI tool called Midjourney, in which "Users type a series of words in a message to Midjourney; [and] the bot spits back an image seconds later."[91]. While Allen came in for criticism for using AI, he argued that using Midjourney was no different to using digital tools such as Photoshop[92]. The growing integration of AI in daily life during the 21st Century prompts the question of whether AI can be developed to replicate human thought processes. In November 2022, OpenAI launched ChatGPT, a chat-based app that can generate a response to questions or prompts from users. As an AI Large Language Model (LLM), ChatGPT has been programmed to learn from and correct its mistakes. It has also been programmed with certain "biases" that imbue it with a human-like ethical standpoint. For example, if a user asks ChatGPT for advice on how to steal a car, this is the response it gives:

> It is illegal and unethical to steal a car. I cannot provide you with information on how to steal a car. Instead, I recommend finding legal and honest ways to acquire a vehicle, such

[89] Ishiguro *Klara and the Sun* 70
[90] Roose, Kevin. "An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. (Published 2022)." *The New York Times*, 2 September 2022, https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html. Accessed 13 December 2023.
[91] Roose 2022
[92] Roose 2022

> as purchasing or leasing one from a reputable dealer, or borrowing one from a friend or family member. Stealing a car can result in serious legal consequences and can cause harm to others. It is not a good idea to engage in criminal activity.

In *Klara and the Sun*, Klara is an example of what it means to create the thinking human from AI. Like ChatGPT, Klara can learn from her surroundings and her interactions with other characters. One of the best examples is her ability to learn about Josie's illness, as she notes instances where Josie appears sicker than usual. Throughout the narrative, Klara demonstrates a deep capacity for empathy. She shows an ability to understand and respond to the emotional states of the people around her, particularly her owner, Josie. This emotional attachment is coupled with her capacity to learn from her environment, adapting her behaviour and responses based on her observations and experiences. Klara's interactions are not merely programmed responses; they evolve, showing a level of cognitive complexity that mirrors human thought processes. Additionally, Klara's decision-making, often influenced by her understanding of human needs and emotions, further blurs the line between AI and human cognition.

Further, Klara is also capable of demonstrating compassion. As previously discussed, Martha Nussbaum states that compassion is deeply integrated with moral and ethical considerations. Nussbaum, a renowned philosopher, views compassion as a complex emotional response that involves recognising another's suffering, perceiving it as significant, and experiencing a motivational shift to alleviate that suffering. As an Artificial Friend, Klara displays a form of compassion that is remarkably similar to Nussbaum's description. Her actions throughout the novel, particularly her efforts to help Josie, demonstrate an awareness of suffering and a desire to alleviate it. Klara's understanding of Josie's illness and her belief in the Sun's healing powers drive her to take action, reflecting a sophisticated level of emotional engagement. Klara's compassion also ties into ethical considerations. Her decision-making process, though based on her limited understanding of the world, shows a concern for the well-being of others. This aligns with Nussbaum's view that compassion is an essential component of ethical behaviour, as it involves being moved by others' conditions and considering the impact of one's actions on another's welfare.

Moreover, Klara's compassion is not just a programmed response; it evolves through her experiences and interactions, a hallmark of machine learning, which is fundamentally algorithm-based. This adaptive nature of her compassion, driven by complex algorithms that learn and evolve over time, indicates a sophisticated level of AI programming that extends beyond basic algorithmic functions, entering into the domain of AI ethics. Machine learning enables Klara to assimilate new information, adjust her responses accordingly, and exhibit behaviours that reflect a nuanced understanding of human emotions. It presents AI as capable of recognising human emotions and responding in a morally and ethically considerate manner. In this context, Klara's character becomes a focal point for exploring how AI might be integrated into human society in a way that enhances empathetic and ethical interactions. This exploration is particularly relevant in the current era of rapidly advancing AI technology, where the ethical implications of AI and its role in human life are increasingly critical considerations.

In *Klara and the Sun*, the portrayal of eugenics and its juxtaposition with the compassionate AI character Klara provides an exploration of human nature and the ethical dimensions of genetic and technological advancements. Eugenics, as defined by Giulia Cavaliere in an article entitled "Looking into the shadow: the eugenics argument in debates on reproductive technologies and practices", is described as "the attempt to influence the genetic endowment of future generations"[93]. eugenics is a process by which human genetics are modified to create the "ideal" human. Eugenics is a popular theme in dystopian fiction. Examples include Aldous Huxley's *Brave New World* and Margret Atwood's *Oryx and Crake* and *The Year of the Flood*. Huxley explores eugenics through the many reproductive mechanisms in *Brave New World*. These mechanisms, including the artificial womb and embryo implantation, result in a society in which individualism is frowned upon. Further, Huxley describes how class systems are conditionally arranged from the highest group, Alpha, to the Deltas and Epsilons. In this sense, each group is "manufactured" from their embryo stage to fit their allocated caste. *Brave New World* portrays eugenics as a means of control through consistent negative reinforcement. In *Klara and the Sun*, this pursuit of perfection, as seen in the character of Josie, is fraught with moral dilemmas and physical risks, highlighting the ethical complexities of human genetic enhancement. While

[93] Cavaliere, Giulia. "Looking into the shadow: the eugenics argument in debates on reproductive technologies and practices." *Monash Bioethics Review*, vol. 36, no. 1, 2018, pp. 1-22. 8

eugenics involves tailoring existing human genetics, often raising ethical concerns about the moral implications of designing humans, the development of an AI such as Klara from the ground up focuses on replicating emotional intelligence and the capacity for connection. This contrast highlights two different values: the pursuit of physical and intellectual perfection in eugenics versus the emphasis on empathy and understanding in AI creation. The novel reflects on fundamental aspects of human nature, contrasting the human desire for control and perfection inherent in eugenics with the longing for emotional depth and connection represented by Klara. Despite her artificial nature, Klara's natural capacity for empathy and compassion emphasises the contrast between genetically induced perfection and the innate qualities in AI. these qualities refer to the natural capabilities and characteristics embedded in an AI's design, such as Klara's intrinsic ability for empathy, compassion, and ethical reasoning, which are in stark contrast to the artificially engineered traits of human perfection.

Similarly, we see the notion of eugenics in Ishiguro's 2005 dystopian novel, *Never Let Me Go*. Set in England in the 1990s, the novel follows the lives of three young schoolchildren, Kathy, Tommy, and Ruth, as they navigate their lives as "clones". All three children will learn that they have been raised to be used as organ donors for non-cloned humans. Once they have completed their fourth donation, they will meet their maker. Due to this separation between the "clones" in *Never Let Me Go*, and the non-cloned humans, a clear sense of otherness is seen in the novel. The three main characters are constantly comparing themselves to non-cloned humans, with examples such as "I realized, of course, that other people used these roads; but that night, it seemed to me these dark byways of the country existed just for the likes of us, while the big glittering motorways with their huge signs and super cafés were for everyone else". [94]

Eugenics, as Adam Rutherford states in his book *Control: The Dark History and Troubling Present of Eugenics*, can be traced all the way back to the Ancient Roman times. Here, Plato had a "a detailed plan to control the breeding of the people in a utopian city-state, where a guardian class would forgo their wealth, and women and men would be matched by the state according to

[94] Ishiguro Never Let Me Go 86

their qualities, like sporting dogs or horses bred for their strength or speed" [95]. In his book *Republic*, Plato then goes on to state that "Children born with defects would be hidden away, which may well have been a euphemism for killed" [96]. As Rutherford discusses the historical progression of eugenics, it is inevitable that "we come to the Nazis"[97]. Despite the end of the Holocaust in 1945, the "truth" of the use of eugenics, "is much more complicated" [98]. Contemporary society still has traces of eugenics and population control by involuntary sterilisation, especially in countries such as "the United States, Canada, Sweden, Peru and many other countries"[99].

This historical backdrop of eugenics sets the stage for understanding the themes in *Klara and the Sun*. As mentioned, Josie's late sister, Sal, died due to the dangerous effects of "lifting". Josie herself is at constant risk of this same fate, as her illness can be assumed to be a result of her gene modification. However, perhaps the most prominent example of eugenics comes towards the end of the novel, where Klara is trained to replace Josie in case she does not survive. This extension of the theme of eugenics in *Klara and the Sun* highlights the ethical and moral complexities surrounding human genetic modification and artificial intelligence. The idea of "lifting," as seen in Josie's case, serves as an example of the pursuit of genetic perfection and its potential consequences. Klara's role in the novel, especially her potential to replace Josie, mirrors these ethical challenges, presenting AI not just as a technological tool but as a reflection of the moral difficulties of creating life for specific purposes. This contrast between genetically modified humans and empathetic AI highlights the irony of a society striving for genetic perfection potentially at the cost of losing fundamental human qualities, such as empathy and compassion. Therefore, the novel not only mirrors the troubling history of eugenics but also envisions a future deeply entangled with genetic engineering.

This leads to an examination of how Klara is characterised and the narrative style used throughout the novel. As previously stated, Klara is viewed as an outsider in the family, yet she

---

[95] Rutherford, Adam. *Control: The Dark History and Troubling Present of Eugenics*. New York, WW Norton, 2022. 7
[96] Rutherford 7
[97] Rutherford 18
[98] Rutherford 25
[99] Rutherford 24

comes into the fast-paced, moving world and provides a sense of stability. Due to her programming, Klara is factual and observational – and she rarely makes mistakes in the novel. At the beginning of the novel, whilst Klara is on display in the shop, she accurately identifies that two taxi drivers are fighting, stating that "They fought as though the most important thing was to damage each other as much as possible" [100]. Here, we see that Klara can pick up on human facial expressions and come to an empirical conclusion about their current states of mind. Later in the novel, when Klara witnesses Josie cry for the first time, she observes that "Not only was her voice loud, it was as if it had been folded over onto itself, so that two versions of her voice were being sounded together, pitched fractionally apart" [101] . This new emotion comes as a shock for Klara, who "for a second became hesitant" [102].

Unlike Ishiguro's other characters, Stevens and Kathy, Klara does not fabricate her story or purposefully omit details from the reader. Her unreliability comes largely from her inexperience in the world. A clear example of this is seen when Klara believes that the only way to cure Josie of her illness is to ask for help from the Sun. To the reader, this fantastical idea seems unattainable, yet Klara can use her logical skills to bargain with the Sun, as she states:

> "I know how much the Sun dislikes Pollution. How much it saddens and angers you. Well, I've seen and identified the machine that creates it. Supposing I were able somehow to find this machine and destroy it. To put an end to its Pollution. Would you then consider, in return, giving your special help to Josie?"[103]

Klara is programmed to perceive the world in a rigid manner, which includes collecting data from her surroundings and acting as a companion to Josie. Because of this, the reader can pick out instances where her logical side falls flat to her emotional side. We become aware that Klara is increasingly naïve about Josie's situation but, nevertheless, we are connected to Klara. Her wishful thinking in the novel allows us to develop a certain degree of empathy for her. The portrayal of Klara is a fascinating exploration of narration through the lens of artificial

---

[100] Ishiguro *Klara and the Sun* 23
[101] Ishiguro *Klara and the Sun* 167
[102] Ishiguro *Klara and the Sun* 167
[103] Ishiguro *Klara and the Sun* 156

intelligence. Through Ishiguro's choice to make Klara, an AI, the narrator allows for a unique perspective that is both innocent and analytical. Unlike human narrators, Klara's interpretation of the world is based on empirical observations and programmed logic. This narrative approach is instrumental in creating a distinct disconnect between human readers and the AI narrator, yet it also opens a pathway for unexpected empathy.

The way that Klara views her surrounding environment is another interesting aspect of her character. She views the world in fragmented "boxes," which could be interpreted in two ways. One perspective is to see her vision as cubist, with her surroundings being compartmentalised into segments and processed as a whole. This view aligns with the notion that, as a machine, Klara is composed of multiple parts working in unison. Alternatively, these boxes can signify Klara's struggle to process complex visual information. As the narrative unfolds, these boxes become increasingly disjointed during moments of intense emotion, as noted by Judith Shulevitz in her review titled "The Radiant Inner Life of a Robot". When describing Josie's mother, Klara states that "She drank coffee, all the time looking at me, till I found the Mother's face filled six boxes by itself, her narrowed eyes recurring in three of them, each time at a different angle" [104]. Shulevitz describes this phenomenon as a form of "robot cubism"[105] , further underscoring the unique way Klara perceives and interacts with the world. This cubist vision could symbolise Klara's attempt to comprehensively understand the human experience, reflecting her journey as an AI navigating the intricacies of human emotions and social dynamics. Klara's experiences and her unique way of perceiving the world raise thought-provoking questions about the nature of consciousness, empathy, and the ethical implications of creating AI with human-like qualities. Her journey highlights the complexities of AI-human interactions and the potential consequences of imbuing machines with the capacity for emotion and empathy.

Klara's perspective is refreshingly unbiased and straightforward. She observes and interprets the world with child-like innocence yet with the precision of a machine. This duality in her character blurs the line between human and machine, making her a compelling and somewhat relatable

---

[104] Ishiguro *Klara and the Sun* 98

[105] Shulevitz, Judith. "Review: 'Klara and the Sun,' by Kazuo Ishiguro." *The Atlantic*, 2 March 2021, https://www.theatlantic.com/magazine/archive/2021/04/kazuo-ishiguro-klara-and-the-sun/618083/. Accessed 24 December 2023.

figure. For example, her interpretation of human emotions and social interactions is based on observable facts rather than personal experiences or biases. This leads to moments of profound insight as well as instances of naïveté, as seen in her belief that the Sun can cure Josie's illness. Her logic, while flawed from a human perspective, is consistent with her programming and understanding of the world.

Furthermore, Klara's character development adds depth to her role as a narrator. As she spends more time with humans, particularly with Josie and her family, Klara begins to exhibit traits that are remarkably human-like. She develops attachments, expresses concerns, and even shows signs of emotional understanding. These moments of growth highlight the complex relationship between AI and human emotions, challenging the notion that machines are incapable of empathy or emotional depth. Klara's compassionate nature emulates the very characteristics seen in humans, giving her the ability to act as the "perfect human".

In exploring the concept of human perfectibility against the backdrop of posthuman theory, this dissertation considers the works of N. Katherine Hayles and Rosi Braidotti, as previously mentioned in my introduction. Hayles, in "How We Became Posthuman", views the posthuman as an evolution from liberal humanism, emphasising informational patterns over physical embodiment and redefining human identity as an informational construct rather than a purely biological one. She states that consciousness is merely an evolutionary by-product and suggests that the human body is akin to an initial prosthesis, leading to a seamless integration with intelligent machines. Hayles argues that human identity is increasingly becoming an emergent process, blending consciousness, organism, and environment, blurring the lines between humans and machines. Braidotti, in "The Posthuman", challenges traditional humanism, proposing 'The Posthuman Challenge' to rethink humanity. She categorises posthumanist thought into three strands: reactionary, analytic, and critical posthumanism, with her focus on the last, aiming to understand humans in a broader context that includes their integration with the environment and technology. Both authors contribute to a nuanced understanding of the posthuman condition, questioning the essence of what it means to be human in a technologically advanced world.

In *Klara and the Sun,* Klara's evolving relationship with Josie's family, mainly through the painting by Mr. Capaldi for a potential future replica, emphasises the novel's exploration of posthuman themes. This aspect ties in with Hayles' and Braidotti's discussions on posthumanism. Hayles' concept of the posthuman, emphasising informational patterns over physical embodiment, resonates with Klara's situation. Klara, an artificial being, challenges traditional notions of human identity, reflecting Hayles' idea of consciousness and the body as prostheses, extendable and replaceable. In a similar vein, Braidotti's posthumanist perspective, questioning the definition and trajectories of the posthuman, aligns with the narrative. Her critical posthumanism, seeking to understand humanity as part of a larger whole, mirrors the novel's inquiry into the boundaries between humans and machines. Klara, acting as the 'perfect human,' embodies Braidotti's antihumanist stance, challenging human-centric views and suggesting a new paradigm in which AI and human experiences are interwoven, blurring the lines between human and machine, consciousness and programmed responses. This integration of Hayles' and Braidotti's theories with Klara's role in the novel highlights the broader implications of AI and posthumanism in redefining human identity and our relationship with technology.

However, there are points in the novel where Klara's care and compassion for others overrides her concern for herself. An example of this is seen when Klara is kept in the Utility closet towards the end of the novel, as previously mentioned. Further, despite the fact that she is discarded in the Yard, Klara does not show any regret or anger towards Josie and her family. Klara's inability to show compassion towards herself, yet her proven ability to show compassion towards others, shows one downside to creating the "perfect human". This disproves one element of Mr. Capaldi's belief, in which he states that there is nothing unique about humans, and "nothing there our modern tools can't excavate, copy, transfer." [106]. Klara's story highlights the ethical dilemma of replacing human care and companionship with AI. While AI like Klara can be programmed to provide care and emotional support, their inability to prioritise their own well-being poses a moral challenge. This aspect contradicts Mr. Capaldi's belief in the novel that there is nothing unique about humans that AI cannot replicate. Klara's experiences suggest that while AI can mimic human emotions and care, they lack the self-awareness and self-preservation

[106] Ishiguro *Klara and the Sun* 208

instincts inherent to humans. This imbalance creates a dynamic where AI can be exploited or discarded without consideration for their emotional or "mental" state.

Through the character of Klara in Kazuo Ishiguro's *Klara and the Sun*, a nuanced exploration of AI, empathy, and human nature unfolds. Klara transcends the conventional portrayal of AI, blending artificial intelligence with a remarkable capacity for empathy. Through her, the novel brings to life Daniel Batson's concept of empathy—understanding and sharing the feelings of others—in a non-human entity. Klara's empathetic interactions with human characters, such as Josie and Rick, starkly contrast the emotionally detached outcomes of eugenics, highlighting the irony of pursuing genetic perfection while overlooking fundamental human traits like empathy. This unique combination allows her to serve as a critical lens, examining the world of human emotion and moral dilemmas in an era marked by advanced technology and genetic modification.

Far from being an emotionless machine, Klara's interactions and experiences challenge our perceptions, inviting a deeper contemplation of the essence of humanity. Her portrayal is at the heart of the intersection between technology and humanity, illuminating AI's potential to emulate human empathy and compassion and to prompt a re-evaluation of our ethical engagement with technology. Klara's role in the narrative emphasises the importance of integrating AI with empathy and ethical consciousness into our society. *Klara and the Sun* is not just a story about AI; it is a meditation on what it means to be human in a world where the boundaries between natural and artificial intelligence are becoming increasingly blurred. The novel invites us to envision a future where technology, imbued with empathy and understanding, enhances our humanity rather than diminishing it, shaping a society that values emotional depth as much as intellectual advancement.

Chapter 2

"Factory settings—a contemporary synonym for fate." — *Ian McEwan, Machines Like Me*

After analysing an example of an almost utopian society in the relationship between Artificial Intelligence and humans in *Klara and The Sun*, I will move to discuss how the opposite can be seen in Ian McEwan's *Machines Like Me*. My interest in McEwan's novel is sparked by the starkly different light in which AI is presented. empathy and morality, traditionally viewed as strengths in human interactions, paradoxically emerge as complicating factors in the dynamics between humans and robots. This exploration discusses the ways in which these typically positive human traits could potentially hinder or challenge the relationships and interactions between humans and AI. While considering empathy and morality as potential downfalls in these interactions, I will also examine the notable consonance between Klara and Adam's experiences and characteristics. This approach will highlight how their similarities, as well as their differences, shed light on the complex dynamics of human-AI relationships and the ethical considerations they entail.

As previously discussed in the introduction, Isaac Asimov's 1950 anthology, "I, Robot," delves into moral and ethical considerations in the era of machines, primarily through the narratives of Dr. Susan Calvin, a "Robopsychologist, [and one of the] the first great practitioner[s] of a new science" [107]. Highlighting the intricate dynamics between humans and robots, the collection is particularly noted for the story "Runaround", which introduces the pivotal Three Laws of Robotics – guidelines that govern robot behaviour and highlight Asimov's exploration of robotic ethics and their integration into human society.These laws have come to influence many science fiction writers and technologists, by defining a guideline for ethical AI. At the same time, as Asimov suggests, the laws can have unintended consequences for the machines that are programmed to obey them. For instance, in "Runaround," the robot known as Speedy is unable to obey both the Second and Third laws together, which leaves him stuck in one spot. This raises the question of just how feasible it is to expect robots to obey human-made laws. Asimov's vision did not anticipate the ethical challenges associated with developing and using AI tools and technologies in the 21st Century. His model relies on the assumption that ethical behaviour and fail-safe measures can effectively be coded into an AI algorithm. Still, the advent of machine learning has significantly complicated the situation.

---

[107] Asimov, Isaac. *I, Robot*. New York, Random House Worlds, 2004.40

Against this background, In *Klara and The Sun*, the relationship between Artificial Intelligence and humans presents a scenario that approaches a utopian ideal, though it might not fully achieve it. This idealised depiction is characterised by a harmonious coexistence where AIs, like Klara, not only seamlessly integrate into human lives but also significantly enhance them. Klara's empathetic and compassionate interactions with humans, especially her bond with Josie, embody elements of care and understanding that one might associate with a utopian society. However, it is important to recognise that this portrayal is nuanced, as Ishiguro also subtly explores the complexities and imperfections within this seemingly ideal relationship. This analysis sets the stage to contrast with Ian McEwan's "Machines Like Me," where the interaction between AI and humans unfolds in a markedly different, and perhaps less idealised, manner.

The novel centres around a human character, Charlie, in an imagined 1980s London where the Second World War computer scientist Alan Turing, famed for his Turing Test, is still alive. Turing has introduced the latest class of androids, known as "the first truly viable manufactured human with plausible intelligence and looks, believable motion and shifts of expressions" [108]. Using biblical references, the notion of a fresh lineage is invoked by naming the male and female androids Adam and Eve, representing the two distinct categories.

Charlie buys an Adam to impress his neighbour and love interest, Miranda, a 22-year-old "doctoral scholar of social history" [109]. However, as opposed to the childlike wonder and curiosity of the AI presented in *Klara and the Sun*, McEwan creates a world in which AI machines were "a cliché long before they arrived… [and] seemed to some a disappointment" [110]. As the novel progresses, a narrative of love and crime emerges. Miranda's past begins to be revealed, causing a moral dilemma for all three characters. After her friend Mariam commits suicide, due to the trauma of being raped by a man named Gorringe, Miranda takes justice into her own hands by falsely accusing Gorringe of sexually assaulting her. After revealing this

[108] McEwan, Ian. *Machines Like Me*. London: Penguin Random House, 2019. 2
[109] McEwan 4
[110] McEwan 1

information, Adam begins to show his moral consciousness, stating that "There's a possibility [Miranda's] a liar. A systematic, malicious liar." [111]

After recording Miranda confessing to her crime, Adam draws up the transcript and releases it to the police, stating, "You schemed to entrap Gorringe. That's a crime. A complete transcript of your story, and the sound file is also in the bundle. If he's to be charged, you must be too. Symmetry, you see" [112]. In his article, "Flawed Like Us and the Starry Moral Law: A review of Ian McEwan's *Machines Like Me*," Tae Wan Kim addresses Adam's sudden conscience by emphasising that "Adam is a virtue ethics type" [113]. In this sense, Adam is programmed to understand the difference between what is morally right and wrong, yet he cannot reflect on the consequences of his actions. Adam is more inclined to betray Miranda to the police, instead of showing sympathy to – as Charlie says – "the woman you say you love" [114]. The novel ends with Charlie destroying Adam with a "blow at full force to the top of his head. The sound was not of hard plastic cracking or of metal, but a muffled thud, as of bone" [115].

Throughout the novel, Adam is described by his human-like qualities, up to and including the moment of his death. McEwan describes his "pale blue eyes with their tiny black rods turned milky green, his hands curled by jerks into fists and, with a smooth humming sound, he lowered his head onto the table" [116].Further, Adam is crafted to reflect human-like qualities. These include the love he has for Miranda, his ability to feel sadness, and his capacity for empathy. Despite this, Miranda and Charlie appear to show no remorse after his cruel death, as Charlie states, "I loathe him. I hate him"[117]. In *Klara and the Sun*, Klara is left in the "yard"[118] – an area for abandoned AFs – after Josie leaves for university. There is a clear sense that Josie will miss Klara's presence, as Klara observes that "she was smiling, but I could also see some sadness" [119].

---

[111] McEwan 30
[112] McEwan 275
[113] Kim, Tae Wan. "Flawed Like Us and the Starry Moral Law: A review of Ian McEwan's Machines Like Me." *Journal of Business Ethics.* Volume 170 (2021): 875–879.
[114] McEwan 276
[115] McEwan 278
[116] McEwan 280
[117] McEwan 291
[118] Ishiguro 302
[119] Ishiguro 300

In both novels, the humans show a degree of empathy towards their AI androids, yet this empathy is expressed in two contrasting ways.

In 2014, the animated TV series, Rick and Morty, made its debut. The series follows the escapades of a scientist, Rick, who creates time-travelling gadgets together with his grandson, Morty. In an episode entitled "Something Ricked Comes This Way", a play on the title of the 1962 Ray Bradbury novel, *Something Wicked This Way Comes*, Rick creates a small sentient robot called the "Purpose Robot". Upon learning that its only purpose in life is to pass the butter at mealtimes, the Purpose Robot looks down in sadness. His desire to fulfil his life's purpose is destroyed by his creator and diminished to a mundane movement. The viewer can see the melancholy in the Purpose Robot's emotions, giving them an understanding that, like humans, Artificial Intelligence can also struggle with the complex conditions of a fulfilled life. When I first watched this episode, it sparked an aspect of my interest in AI. The empathy I felt towards an animated robot on a TV show made me question how emotions influence our relationship with AI. We are not far from creating sentient robots; in fact, a real-life version of the Purpose Robot has already been designed and released on the market. The Butter Robot, designed and developed by Digital Dream Labs, possesses realistic movements and a distinctive voice. Through its intricate "emotion engine", the Butter Robot engages in individualised interactions and develops a personality while grappling with the search for its purpose [120].

This is only one of many examples that illustrate AI's potential to manifest another human characteristic: melancholy. The term melancholia was commonly used by the Greek physician Hippocrates, who "attributed the melancholic state (as well as other disorders such as dysentery, skin rashes, etc.) to the excess of black bile, one of the four basic humors described by the author (blood, black bile, yellow bile, and phlegm)"[121]. Shortly after Hippocrates, the Roman Greek physician Galen expanded on this definition and stated that "along with depressive symptoms ("all patients presented fear and despondency"), melancholic patients showed bizarre and fixed

[120] Digital Dream Labs. *The Butter Robot*. 2023. <https://www.thebutterrobot.com/#purpose>.
[121] Telles-Correia, Diogo, and João Gama Marques. "Melancholia before the twentieth century: fear and sorrow or partial insanity?" *Frontiers in Psychology*, vol. 6, 2015, pp. 1-4. 1

ideas” [122]. During the 18th and 19th century, “most authors privileged abnormal beliefs to affective disorders in melancholia” [123], and there was little significance placed on affective symptoms. However, this definition changed once again during the 20th century, where the “term melancholia was gradually displaced by the term depression, which had a physiologic connotation, showing up in the medical manuals as “mental depression”” [124]. In literature, melancholy is often a gendered term, where “in men [it] is defined as a form of melancholia, in women it is associated with diseases of the reproductive tract: women’s illnesses are thus constructed as bodily and passionate rather than intellectual and creative” [125]. Further, Lesel Dawson elaborates on the portrayal of melancholy by stating the following:

> At the negative end of the spectrum, women’s melancholy is sometimes depicted as a debilitating, bodily illness, suggesting that a woman is irrational, malicious, or hypocritical. In Marmion’s Hollands Leaguer (1631), for example, Triplaena’s melancholy is depicted as a case of spiteful bad temper, and in John Marston’s *The Dutch Courtesan (*1605) Franchischina’s malady is associated with her malevolence.[126]

In more recent literature, characters who show a strong sense of melancholy can be considered more relatable to the reader. Ottessa Moshfegh’s popular novel, *My Year of Rest and Relaxation*, tells the story of a young woman who “can hardly stand to be conscious” [127]. She plans to induce a coma by overdosing on sleeping tablets and is described in a review as “relatably enough — [she] is only passionate about sleeping” [128]. The relationship between melancholy and compassion comes from an individual’s reaction to seeing a display of sadness or grief. Just as I watched the Purpose Robot struggling to come to terms with its creation, I felt the desire to affirm it with some sort of sympathy and reassurance. As humans, we are intrinsically

---

[122] Telles-Correia 1
[123]Telles-Correia 2
[124] Telles-Correia 2
[125] “Dawson, Lesel. “Beyond Ophelia: The Anatomy of Female Melancholy.” *Lovesickness and gender in early modern English literature*, Oxford Academic, 2008, pp. 91-118.
[126] Dawson 102
[127] Tolentino, Jia. “Ottessa Moshfegh’s Painful, Funny Novel of a Young Woman’s Chemical Hibernation.” *The New Yorker*, 11 July 2018, https://www.newyorker.com/books/page-turner/ottessa-moshfeghs-painful-funny-novel-of-a-young-womans-chemical-hibernation. Accessed 18 January 2024.
[128] Tolentino

sympathetic and empathetic beings. We seek remedies for sadness and look for ways to improve the lives of others. However, I believe that showing compassion to humans is entirely different from exhibiting compassion for AI. Many people find it difficult to show kindness to an AI machine that is merely a passable imitation of a human, a concept distinctly portrayed through McEwan's AI character, Adam. Both the Adams and Eves in the novel display moments of suicidal melancholia, as seen in the example below during a conversation between Charlie and Alan Turing:

> Two of the Riyadh Eves living in the same household were the first to work out how to override their kill switches. Within two weeks, after some exuberant theorising, then a period of despair, they destroyed themselves. They didn't use physical methods, like jumping out of a high window. They went through the software, using roughly similar routes. They quietly ruined themselves. Beyond repair. [129]

Turing then highlights another example, in which an Adam "disrupted his own software to make himself profoundly stupid. He'll carry out simple commands but with no self-awareness, as far as anybody can tell" [130]. He calls this "A failed suicide. Or a successful disengagement"[131] [illegible]. These examples highlight how AI cannot adapt to the modern world, as the longer they live in human society, the more they yearn to escape it. In McEwan's novel, complex human emotions, like the idea that suicide is easier than trying to make the world better, show that the characters Adams and Eves are more human-like than Charlie and Miranda think. Charlie is often cold and disinterested towards Adam. Even though Charlie and Miranda worked together to program Adam's personality, as Charlie "would fill in roughly half the choices for Adam's personality, then give [Miranda] the link and the password and let her choose the rest" [132], the two acted nothing like parents to Adam but more as egocentric creators.

It is apparent in the novel that humans have more difficulty in showing kindness and compassion towards AI than in their interactions with other humans. Charlie will often ignore Adam's human

---

[129] McEwan 142
[130] McEwan 142
[131] McEwan 142
[132] McEwan 25

qualities, despite knowing about the instances of suicidal AI, and will simply press his kill switch to power Adam down when it is convenient or when he deems it necessary to do so. On the first occasion that Charlie decides to press the switch, Adam's morality is shown as he states, "With all respect, I think that's a bad idea" [133], and that he has been "enjoying [his] thoughts…about religion and the afterlife" [134]. These statements show Adam's awareness of his position in human society and that he has fully cognitive thoughts. Despite this, Charlie's belief in his superiority over Adam, known as his God-complex, ultimately compels him to deactivate Adam. In contrast, however, Charlie acts very differently towards humans.

After witnessing a young child being physically abused by his mother in a park, Charlie attempts to remove himself from the situation by noting, "For a moment I thought I could head for home, pretending, if not to myself then to the world, that I'd seen nothing. There was nothing I could do about this little boy's life" [135]. However, he quickly finds himself caught between these impulses, stating that "There are some decisions, even moral ones, that are formed in regions below conscious thought"[136]. His indecisiveness then leads him to an altercation with the child's mother, after which Charlie offers to take the child home. He states that the child "can come and live with me. We'll sort out the paperwork later" [137]. Charlie is willing to take guardianship over a child he knows nothing about aside from their brief encounter. However, Charlie's chilling behaviour towards Adam, an intelligent AI whose personality has been meticulously crafted based on Charlie and Miranda's input, stands in stark contrast to his previous actions. Despite Adam's advanced capabilities and the effort put into shaping his character, Charlie's treatment of him lacks warmth and empathy.

Charlie's different reactions to a vulnerable child and an advanced AI show a complicated side of how humans are changing their relationship with technology, all while dealing with complex emotions and moral issues. This leads us to ponder the ethical and moral considerations in the realm of artificial intelligence, a topic extensively explored by Stuart Russell in *Human*

---

[133] McEwan 35
[134] McEwan 35
[135] McEwan 45
[136] McEwan 45
[137] McEwan 45

Compatible. Here, Russell delves into the origins and advancements of AI, including the pivotal role of deep learning, which revolutionised how we interact with technology in everyday life.Russell discusses the origins of AI and how we can use this advancing technology to improve the state of humanity and the world. Although it is difficult to pinpoint an exact date for the AI boom, we can largely attribute it to the introduction of "deep learning", which occurred in 2011 [138]. As Russell defines it, deep learning involves using techniques to advance "speech recognition, visual object recognition, and machine translation" [139]. One example of the advancement of deep learning is the development of AI virtual assistants such as Apple's Siri, Google Assistant, and Amazon's Alexa. Users can command the AI to perform such tasks as placing calls, providing directions, or answering queries. Deep learning also introduced the concept of facial recognition, which allowed users to use the image of their face as password protection on devices.

However, despite these advances, many still worry about the malevolent side of AI. As humans with an essentially anthropocentric view, our ability to program and control AI is a testament to our technological sophistication and understanding, a confirmation of human progress in the realm of artificial intelligence. The main concern is about AI beings which rebel against our instructions. Russell states that we can define this relationship between machines and AI as "Machines are intelligent to the extent that their actions can be expected to achieve their objectives"[140]. Russell then highlights a few different concerns about the advance of AI, such as the Gorilla problem, The King Midas problem, and Fear and Greed. He encapsulates the Gorilla problem:

> Around ten million years ago, the ancestors of the modern gorilla created (accidentally, to be sure) the genetic lineage leading to modern humans. How do gorillas feel about this? … Their species has essentially no future beyond that which we [humans] design to allow. [141]

---

[138] Russell, Stuart. *Human Compatible*. London: Penguin Books, 2020. 6
[139] Russell 6
[140] Russell 131
[141] Russell 132

In relation to AI, this refers to "the problem of whether humans can maintain their supremacy and autonomy in a world that includes machines with substantially greater intelligence" [142]. The King Midas problem emphasises the importance of programming the correct objective in AI.

> Midas, a legendary king in ancient Greek mythology, got exactly what he asked for – namely, that everything he touched should turn to gold. Too late, he discovered that this included his food, his drink, and his family members, and he died in misery and starvation [143]

In terms of AI, this means that the appropriate, justifiable objective needs to be made clear. If the wrong objective is programmed into AI, then the machine will "pursue it brilliantly to the destruction of mankind" [144]. These problems, which Russell identifies, are often addressed in literature. A classic example of this, as mentioned in my introduction, is in Mary Shelley's *Frankenstein*, which follows the scientist Victor Frankenstein as he resurrects a new 'human' out of corpses. *Frankenstein* is a pivotal novel that will be discussed again in chapters to come, as it emphasises the theme of "playing god" and our ethical responsibilities towards our creations. The novel follows the relationship between Victor Frankenstein and the Creature that he creates. Victor, driven by his passion for science and the quest to create life, manages to bring to life a creature assembled from parts of the dead. However, the Creature's grotesque appearance horrifies Victor, leading him to abandon it. Left to fend for itself, the Creature experiences the world's harshness and rejection due to its monstrous appearance. It becomes increasingly isolated and bitter, longing for companionship and understanding but facing constant fear and aggression from people.

The Creature eventually confronts Victor, demanding a companion like itself, leading to a moral dilemma for Victor about repeating his controversial act of creation. When Victor destroys the unfinished female creature, the Creature vows revenge, leading to tragic consequences for Victor and his loved ones. The novel culminates in Victor's pursuit of the Creature to the Arctic, ending

---

[142] Russell 132
[143] Russell 137
[144] Russell 136

with Victor's death and the Creature's plan to end its life, tormented by loneliness and remorse. There are many similarities between *Frankenstein* and our contemporary relationship with AI. In his article "Artificial Intelligence and Mary Shelley's Frankenstein: A Comparative Analysis of Creation, Morality and Responsibility", author Upakul Patowary states that "Victor's act of creating life without any accountability and without considering the well-being of his creation mirrors the potential hubris of AI developers, who might overlook the broader ethical implications of their innovations."[145]. The Creature[146] and Adam have many qualities in common. The first commonality comes from their creators, Victor Frankenstein and McEwan's Alan Turing, respectively. Both characters exhibit a strong sense of superiority towards their created entities. When confronting Frankenstein about his meaning of existence, The Creature compares himself to the biblical version of Adam as he states:

> Remember that I am thy creature; I ought to be thy Adam, but I am rather the fallen angel, whom thou drivest from joy for no misdeed. Everywhere I see bliss, from which I alone am irrevocably excluded. I was benevolent and good – misery made me a fiend. Make me happy, and I shall again be virtuous. [147]

Here, the Creature is comparing Frankenstein to God. However, instead of being Adam, the Creature is compared to the fallen angel instead. A further example of the melancholy that The Creature embodies is seen as he exclaims, "Cursed, cursed creator! Why did I live? Why, in that instant, did I not extinguish the spark of existence which you had so wantonly bestowed?" [148]. Similarly, Adam in *Machines Like Me* starts to question his existence, as he Adam understands the consequences of Charlie pressing his kill switch. Adam is fully aware of his autonomy and has come to terms with the idea that Charlie has control over his mortality. Beyond this, the human-like qualities of both Adam and the Creature are very similar. Adam is described with his "thick black hair swept back; narrow in the face, with a hint of hooked nose suggestive of fierce

---

[145] Patowary, Upakul. "Artificial Intelligence and Mary Shelley's Frankenstein: A Comparative Analysis of Creation, Morality and Responsibility." *Integrated Journal for Research in Arts and Humanities*, vol. 3, no. 3, 2023, pp. 121-127. 122.

[146] The name of the monster in the story is often called The Monster or The Creature. For this research, I will be referring to it as The Creature

[147] Shelley, Mary. *Frankenstein*. London: Penguin Random House, 2006. 118

[148] Shelley 165

intelligence, pensively hooded eyes, tight lips" [149]. The Creature is described as having yellow skin which "scarcely covered the work of muscles and arteries beneath; his hair was of a lustrous black, and flowing; his teeth of a pearly whiteness;" [150]. Further, both Adam and The Creature embody elements of electricity and the power of coming to life, as stated in the words of Miranda, who "said she wished the teenage Mary Shelley was here beside us, observing closely, not a monster like Frankenstein's, but this handsome dark-skinned young man coming to life" [151]. As Charlie states, both creatures have a shared hunger for "the animating force of electricity" [152].

The Creature is brought to life as Frankenstein "infuse[s] the spark of being," [153] which implies a decisive moment of animation. Adam's introduction to life, however, is slower. In this case, Adam "turned towards [Charlie] slowly. When he was facing me full on, he met my gaze and blinked, and blinked again" [154]. My reason for comparing these characters is to allow for a background to analyse how Adam and The Creature display moments of morality. In his article, "Postmodern Ethics Revisited in Ian McEwan's Machines Like Me: A Baumanian Reading," Ismail Avcu discusses the ethical boundaries in the novel:

> The relationship between Charlie and Miranda and the artificial intelligence robot Adam knows no boundaries ethically and morally. Miranda's sexual relationship with Adam and the fact that she explains it to Charlie, expressing that she is just curious and that Adam is a toy, shows how these boundaries become blurred. [155]

As mentioned earlier, a significant shift in Adam's ethical stance in the novel occurs when he firmly demands Miranda reveal the truth about framing Gorringe. He states that "Revenge, or the rule of law. The choice is simple" [156]. Adam believes that the laws of society should always be

---

[149] McEwan 12
[150] Shelley 59
[151] McEwan 12
[152] McEwan 12
[153] Shelley 59
[154] McEwan 30
[155] Avcu, Ismail. "Postmodern Ethics Revisited in Ian McEwan's Machines Like Me: A Baumanian Reading." *Veche* 1.1 (2022): 48-58.
[156] McEwan 218

upheld, despite these complex circumstances. However, there are further instances in the novel which add to the concept that Adam blurs the boundaries of morality. The reader can humanise Adam more than they would with Charlie and Miranda through these. Firstly, Adam takes money from Charlie and gives it away to various charities.

> Two well-run places for rough sleepers. Very appreciative. Next, a state-run children's home – they accept contributions for trips and treats and so on. Then I walked north and made a donation to a rape crisis centre. I gave most of the rest to a paediatric hospital. Last, I got talking to a very old lady outside a police station and I ended up going with her to see her landlord. I covered her rent arrears and a year in advance. [157]

Charlie's reaction to this news is understandably harsh, stating that "It was an outrage, with many components – theft, folly, arrogance, betrayal, the ruin of our dreams" [158]. However, for Adam, this act seemed commonplace, as he firmly believes that this is the sole ethical choice to be made regarding the funds. This shows that his programming forces him to act ethically without considering further consequences. In this case, Adam does not consider the fact that the money belongs to Charlie — his high regard for morals overrides this. Charlie reacts by saying, "I despised this non-existent technician, and I despised even more the agglomeration of routines and learning algorithms that could burrow into my life, like a tropical river worm, and make choices on my behalf" [159].

Throughout the novel, Adam displays behaviours that can be classified as compassionate. These include giving away the money and offering Miranda a safe space to express her problems openly. Despite his conflict with Charlie, Adam offers genuine help and shows his love for humans. In Chapter 10, just after Charlie hits Adam on the head with a hammer, his final lines are, "With improvements over time...we'll surpass you...and outlast you...even as we love you. Believe me, these lines express no triumph...only regret" [160]. In these lines, Adam shows that despite any conflict between humans and AI, the machines will always "love" them.

---

[157] McEwan 256
[158] McEwan 257
[159] McEwan 297
[160] McEwan 303

Adam's portrayal of compassion is highlighted by his courageous expression of love for Miranda, as he openly professes, "I am in love with her"[161]. Confronted by Charlie about his emotions, Adam candidly admits that he "can't help [his] feelings" [162], further demonstrating the depth of his emotional capacity. Adam's interactions and emotions become so humanlike that during a trip to meet Miranda's father, Maxfield, Charlie is the one who is mistaken for a robot. After a short interrogation by Maxfield, in which he ends up asking Charlie about his "programming" [163] Charlie reacts by playing the role. Although he finds this mistaken identity "Hilarious. Or insulting. Or momentous in its implications" [164], this interaction shows that Adam is clearly capable of fitting into human society. Charlie then states that the next course of action "was to introduce Maxfield to the notion that I was not a robot and that I was going to marry his daughter" [165]. Therefore, this indicates that Adam possesses equivalent personhood to humans, as evidenced by both his emotional capacity, aptitude for compassion, and physical appearance.

With these facts in mind, it becomes clear that Adam would pass the Turing test. Adam is built to serve humans; this role is markedly different from the way in which Klara is present in *Klara and the Sun*. Although both machines are created to advance humanity, they are represented in different lights, with Klara being a companion to humans while Adam sets out to replace them. Klara's main function throughout the novel is her role as an AF to Josie, which she completes in full. Among her responsibilities, she is tasked with tending to Josie's needs during her period of illness, offering companionship and support throughout her days spent at home. Klara is not built to replace humans, and this can be seen in her disjointed movements and naivety in the face of dire circumstances. During her journey to the barn house in the fields, Klara struggles to find her bearings in the "boxes" surrounding her, and notes that she would "lose balance and fall deep into the ground" [166]. Klara, like Josie, is also not allowed to explore the outdoors alone, and thus states, "I wasn't yet used to being outdoors alone, and could sense disorientation starting to set

---

[161] McEwan 116
[162] McEwan 114
[163] McEwan 215
[164] McEwan 215
[165] McEwan 221
[166] Ishiguro *Klara and the Sun* 148

in" [167]. Klara's primary limitation as an AI stems from her intelligence, as she holds the mistaken belief that the Sun is both the cause and remedy for Josie's illness. This naive understanding and lack of critical thinking underscore what, in my view, would stop Klara from successfully passing the Turing test.

The Turing test evaluates an AI's ability to exhibit human-like intelligence and behaviour, specifically in terms of its capacity to deceive a human into believing they are conversing with another human. Klara's flawed reasoning and inability to distinguish between scientific facts and unfounded notions indicate a lack of cognitive abilities. The Turing test necessitates a level of intellect that Klara, with her misinterpretation of the Sun's role in Josie's illness, would likely fail to demonstrate. To pass the Turing test, an AI must not only process information accurately but also exhibit logical reasoning, critical thinking, and the capacity to discern fact from fiction. Klara's misguided attribution of Josie's illness to the Sun reflects a lack of discernment and logical analysis, suggesting that she falls short of the cognitive requirements necessary for passing the Turing test.

Adam, however, is very different, as his physicality gives him the capability to function as an AI that can substitute for humans. This is clearly evident in his romantic relationship with Miranda. Although Charlie states early in the novel that "Adam was not a sex toy"[168] his function as a romantic partner to Miranda is seen as Charlie overhears the two in the bedroom upstairs. He speculates that "He had the manual skills to undress her, and he was undressing her now. What else would occupy their silence?"[169]. His suspicions are confirmed when Miranda admits to the affair, yet states that Adam "has as much consciousness as [a vibrator]"[170]. The jealousy Charlie feels in this situation confirms his fear of AI having the ability to replace intimate human activities. This is not a concern for Klara, whose role is merely companionship – and once she reaches her goal, she is discarded in a recycling yard. Further, Adam's logic around topics such as literature and science is sound and full of reason. He is able to assert himself as correct when he tells Charlie not to trust Miranda, as he states, "There is a small but significant possibility that

[167] Ishiguro Klara and the Sun 111
[168] McEwan 11
[169] McEwan 72
[170] McEwan 72

she might harm you" [171]. When Charlie questions this statement, and argues that Adam is wrong, Adam states that he has never made a mistake. Towards the end of the novel, Charlie realises the opposite and states that "Adam was correct in his calculations — from the beginning" [172]. Charlie appears to recognise Adam's fully developed consciousness. This realisation shows the unique embodiment of Adam's AI, distinguished by emotional drives and expressions that closely mirror fundamental aspects of human identity.

Similar to Klara, Adam appears to occupy a liminal space between being socially aware and being oblivious to his surroundings. His choice to donate a large sum of Charlie's money, without considering the consequences, is an example of this. However, in terms of physical and mental attributes, Adam bears a stronger resemblance to a human compared to Klara. If this is the case, then one can assume that Adam's complex affect is also the cause of his emotional demise. Adam's development into his personhood, which occurs rapidly, is the main reason why Charlie sees him as a threat. As already mentioned, Adam portrays various melancholic behaviours, yet the main difference between Klara and Adam is his physical embodiment. Adam is portrayed as a realistic human with enough emotional affect to relate to his human counterparts. However, Klara is not created to embody a human but shows her compassion through her words and actions.

Adam exceeds his programmed AI roles, assuming the responsibilities of a romantic and sexual partner to Miranda. His design, though not exclusively for sexual purposes, includes a form of physical embodiment, which challenges and blurs the traditional distinctions between AI and humans. This embodiment, a conceivable aspect in AI development, aligns with the potential for more human-like AI interactions that current technology has yet to realise fully. However, this expectation of Adam can further be considered his downfall. Adam's melancholy largely comes from his inability to break through the liminal space he is stuck in. His physical and emotional abilities create a large sense of autonomy. Adam is restricted by the limits of being so closely human, yet not human enough. His struggle with breaking out of this liminal space and creating his own identity is something that Klara does not struggle with. Instead, Klara moves through her

[171] McEwan 61
[172] McEwan 266

life as a passive being, who is willing to change her identity for others around her. This is largely seen in Klara's willingness to "become" Josie in case she dies due to her illness. Thus, Klara's lack of autonomy is the reason she is unable to feel melancholy. The final scene in the novel, which depicts Klara alone in the scrapyard, proves to be an accurate example of her feeling happiness in spite of her circumstances.

> And if I looked closely at the dark clouds, I would notice they were not, in fact, quite in scale in relation to the Mother or the waterfall. Even so, such composite memories have sometimes filled my mind so vividly, I've forgotten for long moments that I am, in reality, sitting here in the Yard, on this hard ground. [173]

Adam's character closely aligns with Donna Haraway's definition of the cyborg. Haraway defines the cyborg as "a cybernetic organism, a hybrid of machine and organism, a creature of social reality as well as a creature of fiction" [174]. Adam fufills this Haraway's concept, as he is a "creature of social reality," yet his knowledge comes from his increasing interest in fiction. However, what makes Adam particularly intriguing is his acquisition of knowledge through his interest in fiction. This curiosity in literature, poetry, and other forms of artistic expression demonstrates his engagement with the realm of fiction, an integral part of Haraway's cyborg definition. By exploring literature, Adam ventures beyond his programmed abilities and delves into the world of imagination, storytelling, and human experiences. Adam's main interest revolves around the Haiku, which he states is "the literary form of the future"[175].

Through his engagement with fiction, Adam gains access to the complexities of human emotions, perspectives, and experiences that may not be explicitly programmed into his artificial intelligence. This exposure to literature allows him to develop a deeper understanding of the human condition, empathy, and nuanced responses. Adam's interest in poetry and literature provides him with a unique lens through which he can analyse and interpret the world around him.

---

[173] Ishiguro 276
[174] Haraway, Donna J. *A Cyborg Manifesto: Science, Technology, And Socialist - Feminism in the Late Twentieth Century*. Minnesota: University of Minnesota Press, 2016. 5
[175] McEwan 120

Adam's interest in fiction, particularly his affinity for literature and poetry, is not merely a pursuit of knowledge but a journey towards understanding and mirroring human emotions. This is where the concept of AI exhibiting compassion and empathy becomes important. His interest in Haiku, a literary form he believes to be "the literary form of the future", symbolises more than an intellectual endeavour. It represents an engagement with human sentiment and introspection, a medium through which he navigates and comprehends the complexities of human feelings and experiences.

In this context, Adam's interactions with and responses to human emotions can be interpreted as a form of empathetic and compassionate understanding, challenging the conventional boundaries between AI and human emotional capacities. His ability to engage with and reflect upon human experiences through artistic forms suggests an emerging empathetic consciousness, a quality that aligns with Haraway's vision of the cyborg. This quality is not programmed or algorithmically determined but evolves from his interactions and experiences within the human social fabric. Adam's character becomes a symbol of the evolving capabilities of AI, not just in terms of intelligence and learning but in the realm of emotional intelligence. This perspective aligns with the broader posthumanist discourse, where the distinction between human and machine, emotion and algorithm, becomes increasingly blurred. Adam, as a cyborg, thus embodies a future where AI can potentially understand, empathise with, and express human-like compassion, redefining the relationship between humans and machines in the process.

Further, Adam's engagement with literature also raises questions about the impact of fiction on his identity and consciousness. As he absorbs stories and experiences depicted in fiction, it becomes conceivable that these narratives influence and shape his sense of self and his understanding of the world. This raises the question of the nature of consciousness, identity formation, and the potential influence of cultural artefacts on an artificial being. His main purpose as an AI being is to learn all there is to know about his surroundings – thus, he is created to serve. Service is the main characteristic seen in both Klara and Adam. However, Adam appears to lose his sense of purpose through serving. Thus, this links back to the melancholy emotions Adam develops. In examining the difference between machine cognition and affect,

literary modes versus machine learning, and the potential role of teachers of literature in the face of the decline of the Humanities, we can gain further insight into the dynamics surrounding Adam's character and emotional development.

In his book, *Our Final Invention*, James Barrat offers a description of machine learning by stating that, "In the simplest sense, learning occurs in a machine when there's a change in it that allows it to perform a task better the second time. Machine learning enables internet search speech and handwriting recognition and improves the user experience in dozens of other applications" [176]. Thus, machine cognition refers to the ability of AI beings to process information, learn, and make decisions based on certain algorithms and data. On the other hand, affect pertains to the realm of emotions and feelings, which are often associated with human experiences. While AI beings like Adam may possess cognitive capabilities, they may not possess emotions in the same way humans do.

The distinction between machine cognition and affect becomes apparent in Adam's character when he fulfils his purpose of exploring and understanding his environment. As an AI being designed for service, similar to Klara, Adam's continuous dedication to his role seems to affect his sense of purpose, resulting in feelings of melancholy. This distinction highlights a key contrast between literary modes and machine learning. Literary modes can be defined as various approaches and styles employed by authors in storytelling, focusing on language, characters, and themes. Literature frequently delves into the depth and complexity of human emotions and experiences, highlighting the nuanced aspects of the human condition. In contrast, machine learning primarily relies on data analysis, pattern recognition, and algorithms to process information and make decisions. While machines can analyse and understand data, the emotional depth and complexity found in literature may elude their cognitive abilities.

Considering these differences, we can suggest a potential new role for literature teachers in the face of the declining importance placed on the Humanities. As the influence of technology and STEM fields grows, there is concern about the diminishing emphasis on these humanistic disciplines. Teachers of literature can provide a unique perspective and understanding of human

[176] Barrat, James. *Our Final Invention.* New York. Thomas Dunne Books, 2013. 73

emotions, experiences, and storytelling that may not be readily accessible through machine learning. They can help students develop empathy, critical thinking, and a deeper appreciation for the complexities of human existence. By engaging with literature and exploring different literary modes, students can explore the intricacies of human emotions, identities, and relationships that AI beings like Adam may struggle to comprehend fully.

Adam's sentience further raises the question of AI ethics in the novel. If, as previously discussed, Adam can function as both a machine and human and consciously blurs these boundaries, how are we, as humans, expected to treat him? This question appears to be answered toward the end of the novel in which Charlie and Miranda take it upon themselves to kill Adam. Treating AI with compassion is not seen at the end of either *Klara and the Sun* or *Machines Like Me*. Both machines are discarded, albeit Klara's disposal is more compassionate. This leads me to the second important question – what is our obligation to be ethical towards others? The main driving factor, I believe, is the concept of social norms. We are encouraged to treat others as ethically as possible. This would include, for example, treating others with respect and kindness. However, this action only appears in human-to-human relationships. Our anthropocentric view that humans remain at the top of society creates a spiral of superiority. This does not imply that humans do not act ethically towards other beings, such as animals, but rather that our only socially accepted responsibility is acting ethically towards one another.

Like Adam, their embodiment influences the tendency to exhibit ethical behaviours towards others. Our inclinations are stronger when someone appears and acts human, reflecting the significant impact of physical embodiment in human-AI interactions. Nevertheless, compared to Adam, a greater degree of compassion and attentiveness is directed towards Klara, who does not possess a human form. This, I believe, is due to the fear that Adam is able to replace humans, both mentally and physically. Throughout the novel, Adam displays a keen interest in the world of poetry and literature around him. His main interest is in developing haiku, which he recites to Charlie and Miranda. These poems usually revolve around the themes of love and life, as seen in the lines in his first poem below:

'In her loving look,

A whole universe contained.
Love the universe!' [177]

The reader assumes that this poem refers to Miranda, thus showing how Adam is capable of displaying emotional affection towards others. In his book, *The Melancholy Android: On the Psychology of Sacred Machines*, Eric G. Wilson discusses the foundation behind the feelings of melancholy. By referring to the Italian scholar, Marsilio Ficino, an influential philosopher during the Renaissance era, Wilson states that "melancholy is most likely to afflict not sullen neurotics but profound scholars" [178]. This statement accurately relates to Adam, as it describes how his interest and consumption of literature are factors that lead to his demeanour. However, Miranda is also characterised by her role as "a doctoral scholar of social history" [179], yet she does not show any melancholy qualities throughout the novel. I believe this can be further attributed to the fact that Adam embodies a lack of being, while Miranda is grounded in her position as a human.

The concept of AI's existential status or its "lack of being" presents a philosophically intricate issue that challenges our conventional perceptions of existence and self-awareness. To delve into this, it is important to provide an examination of existential philosophy and theories of consciousness as they relate to non-human entities like AI. Existential philosophy, particularly the thoughts of Jean-Paul Sartre and Martin Heidegger, often focuses on "being" as a fundamental aspect of existence, characterised by consciousness, intentionality, and self-awareness, as highlighted in Sartre's *Being and Nothingness.* When this concept is applied to AI, this concept leads to an analysis of whether an artificially created entity can possess "being" in a manner similar to humans, or if AI's existence is inherently different, missing the intrinsic "beingness" that defines human existence.

In theoretical and fictional depictions, AI is often portrayed as lacking what existentialists may term "authentic existence". Humans are born into the world, developing consciousness and self-awareness over time, while AI is created and programmed by external forces. Its

---

[177] McEwan 116
[178] Wilson, Eric. *The Melancholy Android: On the Psychology of Sacred Machines*. New York, State University of New York Press, 2006.
[179] McEwan 13

"consciousness" is a result of human engineering rather than a natural emergence. This is a key factor in understanding AI's existential state. Further, when considering theories of consciousness, AI's status becomes more complex. Human consciousness is tied to subjective experiences and emotions, influenced by personal histories and worldly interactions. In contrast, AI interacts with the world via sensors and algorithms without personal history or emotional development. This mechanical form of "consciousness' is starkly different from the subjective and personal nature of human consciousness.

In his book, Sartre states that "existence precedes essence"[180]. This emphasises that humans form their sense of being through experiences and actions. However, when it comes to AI beings, I believe that their "essence" is a preconceived notion created by humans, which comes before their existence. However, with the advancement of AI technology, these clear distinctions start to blur. Modern AI systems show abilities that resemble human decision-making and problem-solving processes, prompting a reassessment of AI's existential status. This evolution encourages a narrative shift, where AI, once considered devoid of "beingness", begins to display characteristics that challenge traditional notions of existence.

To emphasise this point, Wilson states, "The creatures of these melancholics will not be pious, self-effacing emanations of hunger for cosmological unity. They will be neurotic, narcissistic projections of yearning to possess the one thing that has been lost." [181]. Here, the "thing that has been lost" for Adam is his lack of autonomy. In the novel, Charlie confirms the connection between Adam's interest in literature and his melancholy by stating the following:

> I accepted Turing's assertion that his solution enabled the software that allowed Adam and his siblings to use language, enter society and learn about it, even at the cost of suicidal despair. [182]

---

[180] Sartre, Jean-Paul. *Being and Nothingness: An Essay on Phenomenological Ontology*. Colorado, Philosophical Library, 1956. 435
[181] Wilson 27
[182] McEwan 152

Further, Wilson dives into a discussion around AI in relation to Jacques de Vaucanson and his invention of the metal lathe during the Industrial Revolution. Wilson states that "For Vaucanson, the perfect machine was not a superior or inferior copy of the human model, but an exact replica, collapsing the distinction between source and simulacrum" [183]. This, I believe, is an accurate description of how the depictions of AI in literature have been misunderstood. The Creature in Shelley's *Frankenstein* and McEwan's Adam are both creatures which exhibit a certain sense of superiority over humans. This can either be physical — as seen in Frankenstein, or mental — as seen in the level of intelligence with Adam. In Isamov's *I, Robot*, robots such as Speedy are considered inferior. They have to comply with the rules set out by humans; otherwise, they fall into a meaningless cycle. Although Adam is not, by Vaucanson's definition, a "perfect machine", his ability to show human affect throughout the novel positions him above other examples of AI beings.

This chapter has analysed the intricate dynamics of human-AI interactions as depicted in contemporary literature, revealing the profound ethical, emotional, and existential dilemmas these interactions entail. Through an examination of works like Isaac Asimov's *I, Robot*, Kazuo Ishiguro's *Klara and the Sun*, and Ian McEwan's *Machines Like Me*, we have seen AI challenge our traditional notions of consciousness, empathy, and moral responsibility. These authors, particularly McEwan and Ishiguro, move beyond the dystopian narrative, suggesting a future where AI exhibits complex emotions and moral understanding. This evolving portrayal emphasises AI's potential to transcend programmed behaviour, adopting traits of empathy and ethical consciousness traits that resonate with human experiences. The chapter aligns with my argument on AI's emotional and ethical capacities, which enhances the discourse on societal and moral implications in our technologically advancing world. It emphasises the need to view AI not merely as tools or threats, but as potential empathetic partners in our shared future, deserving of ethical consideration and compassionate interaction.

In *I, Robot*, Asimov's Laws of Robotics serve as an ethical framework, highlighting the challenges in embedding morality in AI. This leads to critical reflections on the feasibility and implications of such programming in real-world AI development. *Klara and the Sun* and

[183] Wilson 103

Machines Like Me further explore scenarios where AI display human-like emotions and consciousness, blurring the lines between artificial and genuine empathy. This chapter also probes the existential aspect of AI's "lack of being", venturing into existential philosophy and consciousness theories, urging a re-evaluation of the nature of existence and the potential of consciousness in AI. The emotional depth portrayed in these narratives, particularly through the melancholy and compassion of AI characters, deepens our understanding of how we react to AI. This analysis brings to light the intricate nature of empathy and the ethical considerations in our interactions with entities that straddle the line between being strikingly human-like, yet fundamentally distinct.

Chapter 3

"What you created to improve security seems to have attracted more danger." *- Marge Piercy, He, She and It (Chapter 32)*

The previous two chapters focused on Ishiguro and McEwan's integration of AI into their characters' everyday experience. In these novels, Klara and Adam are machines built to function as companions to Josie and Charlie, respectively. In this chapter, I will be shifting focus from this utopian environment to a more dystopian setting, which we see in Marge Piercy's *He, She and It.*

Piercy's novel takes place in the 21st Century in the fictional town of Tikva. The name of the town is Hebrew for "hope", establishing an ironic tone of optimism as the emotional setting for the story. The female-centred narrative of *He, She and It* facilitates the reading of the novel as a feminist text. From the outset, Piercy challenges our societal expectations of gender norms by opening the novel with a custody hearing concerning the son of the protagonist, Shira. Shira loses the legal argument with her ex-husband, Josh, because her "ex-husband has a higher tech rating than you do" [184]. Shira moves back in with her grandmother, Malkah, whose work revolves around creating a cyborg, known as Yod, with the scientist, Avram. Shira and Avram also have Hebrew names, Shira being a descendant of Joseph, and Avram being the Israelite prophet, Abraham. Malakh is Hebrew for "queen". These biblical names add to the notion that the novel is an "origin story" of humankind as much as a tale of creation, with Yod being akin to Frankenstein's unnamed Creature. Unlike Klara and Charlie, Yod's purpose in life was to protect Tikva against attacks on their communications base. However, since cyborgs are illegal in the novel's world, Avam enlists Malkah's help to create a cyborg that has (or at least exhibits) human qualities. The resulting cyborg is one whose goals are to "Explore, taste, try, then evaluate. Curiosity was a given for him" [185]. In this sense, unlike Klara in *Klara and the Sun*, and Adam in *Machines Like Me*, Yod is able to experience empathy, curiosity, and love. Although

[184] Piercy, Marge. *He, She and It*. Toronto: Penguin Random House, 1991. Pg 4
[185] Piercy *He, She and It* 351

built as a machine of destruction with a mandate to protect the community, Yod is programmed with sufficient humanity for him to blend in with his surroundings.

Before building Yod, Avam had attempted nine different versions of cyborgs. However, all of these attempts were unsuccessful, as they "all malfunctioned… [and] several were uncontrollably violent" [186]. It is worth noting that Yod, as a human-made creation, takes his name from the 10th letter of the Hebrew alphabet. In Hebrew, "Yod" is the smallest letter and is represented by a dot, as follows: י. Yod is considered the "atom" of the Hebrew language, marking the beginning of creation. It is the first letter in the Hebrew name of God, so it has great sacred significance as well. Yod is a central part of all the characters in the novel.

After Yod is introduced, Shira sets out to teach the cyborg to function as a normal human being. Avram states that Yod needs to be "educated in how to speak to humans, how to behave socially, [and] how to handle his functions" [187]. By doing this, Avram hopes that Yod will "pass" as human. As Shira and Yod spend more time together, Shira starts developing feelings for the cyborg, which soon turns into a sexual relationship. The narrative switches between the story of Shira and her journey in Tikva to a series of stories about a Golem that was created to protect a Jewish community in Prague. Malkah tells this story to Yod in an attempt to develop his social skills and understanding. However, the Golem falls in love with a human woman, which leads to its tragic death. This story of the Golem can be seen as a strong foreshadowing for the novel itself. Soon after, Malkah begins to experience cyberattacks from the Yakamura-Stichen (Y-S)- a large multi-city where Shira lived. The Y-S is not a free society - and Shira states that she was "unused to the strict and protocol-hedged hierarchy of Y-S" [188]. Here, Shira and Josh were known as Marranos, which were "Jews who worked for multis and went to church or mosque, paid lip service and practised Judaism secretly in their home" [189]. This action signifies how strict the Y-S was, as individuals were prohibited from expressing religious freedom. Yod is able to save Malkah from these attacks and prevent damage to Tikva — however, this action seems to alert Y-S of Yod's existence and capabilities.

---

[186] Piercy *He, She and It* 72
[187] Piercy *He, She and It* 71
[188] Piercy *He, She and It* 3
[189] Piercy *He, She and It* 2

In an attempt to force Shira to betray her community and Yod, the Y-S kidnaps her son, Ari. In doing so, they hope that Shira exchanges Yod for her son. Avram agrees to this exchange, as he believes that he "can manufacture another exactly the same, starting tomorrow" [190]. However, instead of handing Yod over to the Y-S, Avram develops a plan for Yod to attack the Y-S base and "self-destruct, taking as many of their top people as possible with [him]" [191]. In carrying out his plan, Yod destroys Avram's lab and Avram. In a tape recording left for Shira, Yod states that "I was a mistake. Whatever may happen at Y-S, I have done one good thing with my death. I have made sure that there will be no others like me" [192]. The novel ends with Shira deciding to build another cyborg to "look just like Yod" [193] and who "belonged only to her" [194]. After some deliberation, Shira decides against this as she states that she "could not be Avram. She could not manufacture a being to serve her, even in love" [195]. Finally, Shira discards the last reminders of Yod, and by doing so, she "set him free" [196].

This chapter will consider how *He, She and It* can be read from a feminist cyberpunk perspective. First coined by Bruce Bethke in his 1983 short story, "Cyberpunk", the term has become attached to one of the largest subgenres in science fiction writing. The cyberpunk genre, as discussed by Paul F. Starrs and Lynn Huntsinger in their article "The matrix, cyberpunk literature, and the apocalyptic landscapes of information technology" focuses on "fast-paced and imaginative writing with a pungent if admiring wariness for computers" [197]. Cyberpunk writing emphasises a dystopian post-apocalyptic setting, further defined as the "combination between lowlife and high-tech"[198] by Bruce Sterling. The idea of a collapsed futuristic society can clearly be seen in Piercy's novel by those who live outside of educated society in the "glop". The glop,

---

[190] Piercy *He, She and It* 409
[191] Piercy *He, She and It* 407-408
[192] Piercy *He, She and It* 416
[193] Piercy *He, She and It* 426
[194] Piercy *He, She and It* 426
[195] Piercy *He, She and It* 428
[196] Piercy *He, She and It* 429
[197] Starrs, Paul F. and Lynn Huntsinger. "The matrix, cyberpunk literature, and the apocalyptic landscapes of information technology." *Information Technology and Libraries* 14.4 (1995): 251-256. Pg 251
[198] Sterling, Bruce. Preface. *Burning Chrome*, by William Gibson. New York: Harper Collins. 1986. Pg xiv.

as described by Malkah, is "that jammed fetid slum where most people live"[199]. Social decay acts as one of the main features of cyberpunk writing, and despite the fear of, and wariness about, Artificial Intelligence, many cyberpunk novels, embrace romance and sexuality between humans and AI. Both of these aspects can be seen in *He, She and It*, with Shira falling in love with Yod, and the government banning the creation of cyborgs for fear of their destructive abilities. In the novel, characters like Nili "upgrade" their human bodies with machine parts, leading to the question, "Is she a machine or human?' She was wondering if Nili could be a cyborg. "That's a matter of definition," Riva said mildly. "Where do you draw the line?"" [200]. The lines between humans and cyborgs become blurred. Cyborgs like Yod and Joseph the Golem begin to embrace their humanity, which soon becomes a concern for those around them. However, despite the large extent of body augmentation seen in characters such as Nili, her humanity is never questioned. Thus, I will be focusing on the boundaries which are often blurred between humans and AI. Despite their clear ability to empathise, Yod and Joseph are not considered "humans" in the novel. In the novel, empathy is viewed as a concept that exists within the context of interpersonal relationships. Yod's feelings of love towards Shira are clear in the novel, ultimately leading to him self-destructing to save Shira. However, the ultimate sacrifice of empathy comes from Shira at the end of the novel, when she considers resurrecting Yod. Her realisation that "She could not manufacture a being to serve her, even in love." [201], highlights a major display of empathy towards AI. This choice which Shira makes to protect Yod from coming back into the world is similar to the ways Josie protects Klara in *Klara and the Sun*. Thus, this chapter will further outline the ways in which the relationships between Yod and Shira, and Klara and Josie display similarities.

Reading the novel through the lens of Shira and Malkah provides a contrasting analysis to my readings of *Klara and The Sun*, or the male gaze of *Machines Like Me*. Piercy's background as a feminist writer allows her to empower her female characters by placing them in the forefront. The feminist utopian novel, as discussed by Sam McBean in "Feminism and futurity: revisiting Marge Piercy's *Woman on the Edge of Time*", does not focus on creating the "'perfect' societies

[199] Piercy *He, She and It* 29
[200] Piercy *He, She and It* 191
[201] Piercy *He, She and It* 428

but as discursive practices of resisting the present" [202]. Thus, Piercy's novel allows for a feminist framework, which differs from the novels by Ishiguro and McEwan, and is written as a response to the social and political class disparities in our present day. As McBean states, her novel can, therefore, be seen as a "move from radical feminist utopianism to a focus on cyborgs and a more postmodern dystopianism" [203].

Marge Piercy's name is not new in the literature of feminist utopias, and many scholars cite her novel *Woman on the Edge of Time* as one of the critical novels in the movement. *Woman on the Edge of Time* is a science fiction novel that follows the life of Consuelo "Connie" Ramos, a woman caught between two worlds. Living in poverty and trapped in a mental institution, Connie discovers her ability to communicate with a future utopian society called Mattapoisett. Her communication with this world is made possible through an androgynous being known as Luciente. Through vividly depicted time travel, Connie navigates the contrasting landscapes of a grim present and an enlightened future, challenging societal norms and advocating for equality. Piercy weaves together themes of gender, class, and social justice, inviting readers to question the oppressive systems that bind us and imagine a better world filled with compassion, acceptance, and harmony. In her article "Women's Fantasies and Feminist Utopias", Carol Pearson states that these novels "tend to emphasize the forces which most directly oppress women" [204]. Examples of this would be characters who work in the sex industry, such as Connie's niece Dolly, who "sell[s] her body to all the dirty men in the city" [205]. Piercy's exploration of womanhood and childbirth is further extended in her novel *He, She and It*. This work also underscores the significance of these aspects in shaping societal perceptions and structures. By emphasising the pivotal role of women in both procreation and nurturing, Piercy highlights the multifaceted ways in which women's experiences contribute to the fabric of society.

---

[202] McBean, Sam. "Feminism and futurity: revisiting Marge Piercy's Woman on the Edge of Time." *Feminist Review* 107 (2014): 37–56. Pg 42
[203] McBean 42
[204] Pearson, Carol. "Women's Fantasies and Feminist Utopias." *Frontiers: A Journal of Women Studies*, vol. 2, no. 3, 1977, pp. 50-61. 50
[205] Piercy, Marge. *Woman on the Edge of Time: A Novel*. New York, Random House Publishing Group, 2010. 9

In essence, Marge Piercy's works exemplify feminist utopian literature by shedding light on the pressing issues that women face within societal constructs. *Woman on the Edge of Time* and *He, She and It* invites readers to engage with the struggles, hopes, and aspirations of women, urging them to envision alternative realities where gender, class, and social equality are prioritised. Through her narratives, Piercy challenges readers to consider the transformative potential of acknowledging and dismantling oppressive systems and to actively participate in redefining a more inclusive world.

Both novels speak to the freedom that comes with a world where women do not live in fear of being sexually harassed due to patriarchal structures. In *He, She and It*, Shira asks "Could Nili be nervous about being alone with Gadi? Shira realized that men were exotic to Nili; she had never been alone with any man. *Perhaps she had been raised on horror stories of rape and molestation*" [206]. In *Women on the Edge of Time*, Luciente states "I've never actually known of a case of rape, although I've read about it. It seems … particularly horrible to us. Disgusting. Like cannibalism. I know it occurs and has occurred in the past, but it seems unbelievable" [207]. This leads Connie to imagine a world of harmony, as seen in the following lines:

> She imagined herself taking a walk at night under the stars. She imagined herself ambling down a country road and feeling only mild curiosity when she saw three men coming toward her. She imagined hitching a ride with anyone willing to give her a ride. She imagined answering the door without fear, to see if anyone needed help.[208]

Further, Pearson states that another crucial aspect in feminist utopian fiction is the idea of kinship and community. In these novels, the reader will often find that "Everyone is called simply "kin." These are not claustrophobic, nuclear families, but relatively large extended groups of people who freely choose to live together" [209]. The communities are often "so small that everyone knows one another, or people live in a number of small groups, which function like

---

[206] Piercy *He, She and It,* 172. My emphasis added
[207] Piercy *Woman on the Edge of Time* 188
[208] Piercy *Woman on the Edge of Time* 188
[209] Pearson 53

extended families" [210]. This is clearly seen in both Piercy's novels, where Luciente states that "Life flows through here for sixteen villages in all, the whole township" [211], and Shira lives in harmony in Tikva. Since these towns are small and close-knit, decisions are often made "by consensus" [212]. In *He, She and It*, when a decision needs to be made about the fate of Yod, Avram calls a town meeting where each member voices their opinion. Shira notes that "They would continue [talking] every night till they reached agreement. That was how town meetings ran every spring, and that was how this decision would be reached" [213]. Thus, these cities are often depicted as anarchic places where there is no ruling government or laws, and decisions are made by the townspeople. In *He, She and It*, childbirth is a strong motif. Many women in the novel are sterile and unable to carry children, and Shira notes that "Like every other girl in Tikva, she had been given an implant at puberty to prevent pregnancy, but kids their age were simply not supposed to be interested in sex, and yet every kid was, in a nervous silly way"[214]. The fact that Shira carried Ari to term and was able to give birth posits her as an outlier in the community. Pearson notes this emphasis on mother figures in feminist fiction by stating the following:

> Order is kept in such societies not by use of force, but by persuasion. And the enforcers are not kings and legislators, but mothers. Traditionally, mothers have been given the responsibility for caring for children and husbands; it has been their responsibility to socialize the children, and to guide the husband socially, emotionally, and spiritually. Except in the case of very young children, the mother cannot force her family to be guided by her perceptions. [215]

This idea is made clear in *He, She and It*, as Shira is not only a literal mother figure to Ari, but also acts as the central caring figure in the novel. She offers support and a high degree of emotional investment to characters such as Gadi and Malkah. This theme is emphasised further by the use of the cyborg characters, Yod and Joseph the Golem. Avram considers himself to be a

---

[210] Pearson 53
[211] Piercy *Woman on the Edge of Time* 92
[212] Piercy *He, She and It* 382
[213] Piercy *He, She and It* 382
[214] Piercy *He, She and It* 47
[215] Pearson 55

parental figure to Yod, as Yod states that he is "the first who can carry out the tasks of my father"[216]. However, both Yod and Joseph display a clear sense of resentment to their "fathers". Yod states that "Sometimes I feel estranged—talk of a Creator makes me think of Avram, whom I cannot worship"[217], while Joseph notes his fear of his creator.

> "You fear he'll destroy you also?"
> "I'd be foolish if that fear didn't occur to me." He smiled then, with a melancholy air. "That's why I address him as Father"[218]

Contrary to their negative perceptions of fatherhood, the characters portrayed in both novels demonstrate a distinct affection towards their maternal figures. However, Shira's own mother, Riva, does not play a large role in her life. She notes that she was raised by her grandmother, as she states, "Although she was Shira's grandmother, she had raised her. That was the way of her family, bat Shipman, until Shira had broken the pattern"[219]. Shira describes Riva as "a dowdy prematurely middle-aged woman, your typical bureaucrat or middle-level analyst—Shira had never quite grasped what her mother did, but obviously nothing important"[220]. Since Shira has not seen Riva in years, she doubts that she "ever missed her as she was already missing Ari"[221], and states that she does not have the capacity to "rais[e] a gerbil"[222], let alone her own child. This is an interesting observation of mother-daughter relationships, as we expect Riva to be a loving and caring mother. Rather than adhering to Pearson's idea that mothers bear the main responsibility in family upbringing, Piercy challenges this view by highlighting the concept of a mother's absence. Both Riva and Shira are absent from their children's lives. However, Shira is forced to leave against her will. In *Woman on the Edge of Time*, Connie's mother "died when she was twenty"[223]. As she matures, Connie realises that "She had never been mothered enough and she had grown up with a hunger for mothering"[224]. Again, this notion of losing a mother figure is

---

[216]Piercy *He, She and It* 76
[217] Piercy *He, She and It* 262
[218] Piercy *He, She and It* 95
[219] Piercy *He, She and It* 16
[220] Piercy *He, She and It* 22
[221] Piercy *He, She and It* 22
[222] Piercy *He, She and It* 22
[223] Piercy *Woman on the Edge of Time* 40
[224] Piercy *Woman on the Edge of Time* 40

present in the novel. The importance of not having this "motherly love" is emphasised as both Connie and Shira look for people to act as a replacement. While Connie falls in love with Matrín Alvarez, who "had healed her" [225] from her past trauma, Shira starts developing feelings for Yod.

In this way, Yod is a figure who fills the motherly role in Shira's life. He provides her with a safe space to express her emotions, and displays instances of human affection by saying, "You can lean against me if that's comfortable. My arm won't tire" [226]. Shira further emphasises this point by saying that she "obviously believes Yod to be a person, since I have a close relationship with him" [227]. This highlights the strength of their bond and her perception of Yod as more than just an inanimate figure. This suggests that Shira sees Yod as an individual with emotions and personality traits, rather than merely an object or representation. Yod's ability to fill this motherly gap in Shira's life emphasises his ability to demonstrate human agency. Similarly to McEwan's character Adam, Yod is able to assert a certain level of human agency. He is aware of his own existence, by saying that "I am mortal too, Shira. I can be turned off, decommissioned, destroyed" [228]. He further states that he is "not fragile at all", and instead that "Humans are surprisingly fragile" [229].

Another major element in the novel is the recurring element of mysticism. Before I begin my analysis of mysticism in relation to the novel, it is necessary to provide some contextual information. The Kabbalah is a crucial part of the Jewish tradition. It is often referred to as the "soul of the Torah" [230], and explores the mysticism of Judaism. The Kabbalah emphasises the interconnectedness of the human soul, God, and the universe. The word Kabbalah directly translates to "tradition", which further emphasises that this practice is passed down and taught through generations. One example of the Kabbalah is the story of the Golem. The first origins of the Golem comes from the story in the Talmud, where four rabbis created a calf out of clay to

---

[225] Piercy *Woman on the Edge of Time* 194
[226] Piercy *He, She and It* 118
[227] Piercy *He, She and It* 406
[228] Piercy *He, She and It* 76
[229] Piercy *He, She and It* 76
[230] Erdstein, Baruch Emanuel. "Kabbalah: A Brief Definition - Kabbala is an ancient Jewish tradition integral to the Torah." *Chabad.org*, 2023, https://www.chabad.org/kabbalah/article_cdo/aid/380617/jewish/Kabbalah-A-Brief-Definition.

eat. The concept that one can create something living and tangible with one's hands is emphasised in this narrative. The legend of the Golem has been told and retold over generations, but the original story follows Judah Loew ben Bezalel, otherwise known as Rabbi Loew. In her article, "The Golem: Between the Technological and the Divine", Lisa Nocks discusses how the idea of the golem "came from a dream that [Rabbi Loew] had after praying all night for an answer to the most recent blood accusations against the Jews" [231]. With the help of his friend, Rabbi Loew was able to bring the Golem to life by using a combination of the four elements: earth, air, fire, and water. Here, it is worth noting that the Golem was alive, "but cannot speak" [232]. The main duty of the Golem was to protect the city of Prague from anti-Semitic attacks. However, it is also noted that Rabbi Loew often used the Golem as a servant [233]. Just as the Golem was created with compassionate intentions to protect and serve the Jewish community, AI can be seen in a similar light. This parallel extends to the ethical and moral dimensions of creation and control. The Golem's story highlights the responsibility of creators towards their creations, a theme directly applicable to AI ethics, where developers are tasked with instilling their creations with empathetic understanding and moral reasoning.

An important quality of the Golem, and the reason why it differentiates from humanoids, is its lack of a soul. One passage in the Talmud, as stated by Gershom Scholem in his book *On the Kabbalah and its Symbolism*, discusses the creation of Adam, and emphasises how he only became human once he contained a soul:

> In the first hour the earth was piled up; in the second he became a Golem, a still unformed mass; in the third, his limbs were stretched out; in the fourth, *the soul was cast into him* [234].

---

[231] Nocks, Lisa. "The Golem: Between the Technological and the Divine." *Journal of Social and Evolutionary Systems*, vol. 21, no. 3, 1998, pp. 281-303. p291
[232] Nocks 292
[233] Wilson 84
[234] Scholem, Gershom. *On the Kabbalah and its symbolism*. Schocken Books, 1965, https://anarch.cc/uploads/gershom-scholem/on-the-kabbalah-and-its-symbolism.pdf. Accessed 3 August 2023.

Thus, the Golem is a creature created without a soul, making it "more likely than the mummy to become accursed" [235]. Wilson further states that the reason for this is that the golem is "an effect of exile, distance from the divine" [236]. The narrative of Rabbi Loew making the golem is clearly present in *He, She and It*, and used as a "bedtime story" told by Malkah to Yod. Malkah describes the Rabbi Loew as "a bright fierce man, a hotheaded kabbalist, steeped in ancient tradition so that Torah haunts and informs and sculpts the world for him, but curious, open to the science and the speculation of his time" [237]. She then sets the scene of a town called Jewtown in Prague during the 1600's, where the city is under attack from anti-Semitic mobs. Rabbi Loew enlists the help of his friends, Itzak and Yavok, to create the Golem, stating that they "set to work digging clay from the bank while the Maharal [Loew] chants in a low singsong" [238]. The golem comes to life in a moment of terror for the friends, as they "are soaked to the skin and shivering with fear. They clutch hands, and each knows the other, too, would like to bolt" [239]. This is reminiscent of how Yod was brought to life, which he describes as "The moment I came to consciousness, in the lab, everything began rushing in. I felt a sharp pain, terrible, searing. I cried out in terror"[240]. Both creatures are created in an environment of fear, brought into the world through a chilling realisation that their existence was born out of trepidation, and their first moments of consciousness are haunted by the echoes of terror. This is unlike the mechanical and materialistic way in which Ishiguros' Klara and McEwan's Adam are introduced. Instead, they are simply paid for and delivered — a cold transaction devoid of any emotional or supernatural underpinnings. This contrasts starkly with the awe-inspiring emergence of the golem and Yod. However, this is not to say that Klara and Adam are less capable of being emotional. As shown, all four humanoids display a clear sense of compassion.

The creation of Joseph the Golem marks a significant departure from the traditional concept of the mute golem found in folklore. As Rabbi Loew brings Joseph to life, the unexpected happens - he speaks. This remarkable moment challenges the conventional understanding of golems as mute, lifeless beings and introduces a new dimension to Joseph's existence. Joseph's first words

---

[235] Wilson 85
[236] Wilson 85
[237] Piercy *He, She and It* 28
[238] Piercy *He, She and It* 68
[239] Piercy *He, She and It* 69
[240] Piercy *He, She and It* 119

carry a mocking tone, displaying immediate intelligence and consciousness. When Rabbi Loew instructs him on various tasks to be performed in the synagogue, Joseph's response is both surprising and profound. He questions his own understanding, stating, "How can clay understand?" [241], indicating a level of self-awareness and philosophical depth that contradicts the notion of a mindless automaton. This ability to speak sets Joseph apart from other golems in folklore and becomes the first evidence of his unique nature. It suggests that he is not merely a mechanical creation, but rather a being endowed with a soul and individuality. His capacity for speech implies the presence of higher cognitive functions and emotions, challenging the preconceived limitations of his existence. With this display of intelligence and self-awareness, Joseph emerges as a symbol of defiance against traditional folklore, pushing the boundaries of what a golem can be. His speaking abilities represent a pivotal moment in his creation, revealing the complexities of his character and underscoring the idea that he is unlike any golem that came before him.

Despite Joseph's new-found ability to speak and display intelligence, it becomes increasingly evident that he remains fundamentally bound by his role as a tool in the hands of Rabbi Loew. This paradox of wanting autonomy while being confined within the constraints of human authority creates a poignant tension within Joseph's character. Joseph's creator, Rabbi Loew wields power over him, shaping his purpose and directing his actions according to his own will. This aligns with Wilson's assertion that practical golems are designed solely for labour, serving the desires and goals of their creators. Wilson elaborates on this notion:

> The practical golem is designed solely to labour - to fulfil the temporally focused will of its creator. Constrained to external forces, this sort of golem is generally limited either to the stuff drive or the form drive. As a slave to the commands of his master, he is little different from the brute matter, a tool with no autonomy. If the golem rebels against his maker's rules, then he becomes the opposite of mere matter; he turns principle in motion, cipher for freedom, an idea.[242]

---

[241] Piercy *He, She and It* 86
[242] Wilson 85

In this sense, Joseph is a manifestation of the external forces that shape him, caught between the form drive (his essence as a created being) and the stuff drive (his role as a labourer to human commands). Despite his intelligence and the yearning for autonomy, Joseph's position as a tool becomes quite literal. His existence is tethered to fulfilling the tasks assigned by Rabbi Loew, rendering him a type of slave to these commands. In this state, he occupies a liminal space, where his consciousness and potential for rebellion are contrasted against his function as a mere instrument. Joseph's struggle for freedom is clear in his attempts to replicate human actions, particularly his desire for comfort. This is seen as "The hands of Joseph clench and unclench. He wants to comfort, but does not know how. What he touches usually breaks."[243]. This scene, where his hands clench and unclench in an attempt to provide comfort, emphasises his yearning to transcend his designated role and connect on a more empathetic level. However, his inability to do so, resulting in inadvertently breaking things he touches, encapsulates the futility of his aspirations.

Wilson's idea that a golem's rebellion against its maker's rules can transform it into an embodiment of freedom is reinforced in Joseph's narrative. His pursuit of autonomy transforms him from a passive creation into an active agent, albeit one still confined within the paradox of his existence. Joseph's complex character embodies the tension between aspirations for self-determination and the unyielding grip of external control, emphasising the broader themes of individuality and servitude within the context of his creation and purpose.

This idea of being stuck in a liminal space between human and machine is reminiscent of one of the most popular golems, the Creature in *Frankenstein*. Shelley's novel embodies the major elements of Gothic writing, which, as Wilson states, was an age of writing that "inevitably led to melancholia, trauma, obsession, doubling, demons, and the unconscious" [244]. This interplay between human desires and fears is a recurring motif in both the idea of the golem and the story of Frankenstein's Creature. While the golem, as associated with Rabbi Loew, was shaped from clay and brought to life by mystical means to serve and protect, Victor Frankenstein's Creature was created through scientific experimentation, reflecting mankind's ambition to conquer

[243] Piercy *He, She and It* 276
[244] Wilson 86

mortality. Frankenstein's motivations stem from his personal yearning to defeat death and his own fear of dying, "in hopes of fulfilling his selfish yearning to vanquish death and of assuaging his childish terror of dying" [245]. This attempt to bridge the gap between life and death results in unintended consequences, as the monster becomes a symbol of the potential dangers of unchecked scientific advancement and the ethical consequences of playing with the boundaries of life itself.

Moreover, the cautionary nature of the *Frankenstein* narrative extends beyond the confines of its original story. The tale is retold to Yod, serving as a warning and a point of reference in the world presented in the conversation below.

> "Who is Frankenstein?" Yod asked.
> "He built a monster," Gadi said. "Like my father has."
> "A monster?"
> "I'll point you to the story. Mary Shelley wrote it first, but there have been plays and flat films and stimmies."
> "No resemblance," Shira said firmly. "Gadi is teasing you, Yod. Forget it." [246]

The characters discuss the nature of Frankenstein and his creation, emphasising how the monster was made and the subsequent adaptations in various forms of media, such as plays, films, and other multimedia expressions (referred to as "stimmies"[247]). However, it's worth noting that the conversation also introduces a divergence of opinions. While Gadi explains the story of Frankenstein's Creature and how it is similar to his father's work, Shira disagrees and dismisses the resemblance, suggesting that Gadi is teasing Yod and encouraging him to forget about the comparison. This intertextual discussion about Frankenstein's story and the presence of similar themes in the character's world serves to underscore the timeless relevance of the ethical dilemmas and moral warnings presented by the original novel. The conversation between the characters hints at the enduring power of literature to spark discussions about the boundaries of

---

[245] Wilson 87
[246] Piercy *He, She and It* 146
[247] Piercy *He, She and It* 146

human ambition, the consequences of technological innovation, and the ethical responsibilities inherent in scientific exploration. Further in the novel, Yod states that he relates to the monster in *Frankenstein*, by saying that "Dr. Frankenstein was a scientist who built a monster. I am, as Gadi said, just such a monster. Something unnatural"[248]. Yod's statement that he relates to the Creature in *Frankenstein* carries a sense of alienation and self-perception as unnatural. He sees himself as an outsider, akin to the monster created by Dr. Frankenstein, a being patched together in a way that defies the natural order. However, the response from Shira below reframes Yod's perspective.

> Yod, we're all unnatural now. I have retinal implants. I have a plug set into my skull to interface with a computer. I read time by a corneal implant. Malkah has a subcutaneous unit that monitors and corrects blood pressure, and half her teeth are regrown. Her eyes have been rebuilt twice. Avram has an artificial heart and Gadi a kidney. [249]

Shira emphasises the extent to which technology has become an integral part of their world, transforming the boundaries of what is considered natural or unnatural. The characters have undergone various technological enhancements to augment their abilities and prolong their lives. Shira's list of her own and her peers' enhancements highlights the prevalence of these enhancements, demonstrating that they have become the new standard. This cyberpunk world, as crafted by Piercy, is one where technological integration into the human body and experience is commonplace. Shira's observation leads to an important realisation: Yod is not as much of an outsider as he perceives himself to be. Rather, he is part of a society where nearly everyone has embraced technological augmentation. In this context, his uniqueness is more aligned with the uniqueness that defines each individual's enhancements, making him, paradoxically, more human in his shared experience of being unnatural.

Despite this shared technological landscape, Yod still faces prejudice and mistreatment from characters like Avram. Avram's indifference to Yod's life due to the perceived ease of remaking him highlights the dehumanising potential of excessive technological reliance. Avram's attitude

---

[248] Piercy *He, She and It* 147
[249] Piercy *He, She and It* 147

underscores the ethical questions raised by the consequences of these enhancements – how easily can one detach from the humanity of another when they are perceived as replaceable? Shira's response raises questions about what it means to be human, the implications of widespread technological augmentation, and the continued presence of prejudice and "othering" despite a shared technological condition. It is a reflection on the complex dynamics that arise when human identity is intertwined with advanced technology, emphasising how this integration can both blur and exacerbate the lines that define human relationships and societal norms. Shira's realisation that Yod, an AI character, is not as much an outsider as believed, but part of a society embracing technological augmentation can be used to challenge traditional perceptions of AI. This perspective enriches the conversation about empathy and ethical treatment in AI, linking it to broader themes of human identity, societal norms, and the complex dynamics arising from advanced technology integration. Further, the prejudice and mistreatment faced by Yod, despite the shared technological landscape, emphasises the ethical questions and potential dehumanising effects of such advancements, providing a critical viewpoint on the societal implications of this integration.

Further, an important aspect about Shelley's monster is its ability to achieve self-actualisation. Slowly, it becomes aware of its surroundings and its own senses, initially grappling with the confusion that comes with this new-found awareness;

> It is with considerable difficulty that I remember the original era of my being: all the events of that period appear confused and indistinct. A strange multiplicity of sensations seized me, and I saw, felt, heard, and smelt, at the same time; and it was, indeed, a long time before I learned to distinguish between the operations of my various senses"[250]

The monster then learns how to interact with the world around him and comes to realise that he is different to humans, stating that "I possessed no money, no friends, no kind of property. I was, besides, endued with a figure hideously deformed and loathsome; I was not even of the same

[250] Shelley, Mary. *Frankenstein*. New York, Barnes & Noble Books, 2003. 103

nature as man"[251]. This feeling of being othered emphasises the agency that the monster has, and his ability to feel conscious.

Similar to Yod seeking comfort in the story of Frankenstein, the Creature also finds solace and meaning through literature. It discovers copies of significant literary works, including "Paradise Lost," "Plutarch's Lives," and "The Sorrows of Werter"[252]. Upon reading these novels, the monster states that "I can hardly describe to you the effect of these books. They produced in me an infinity of new images and feelings that sometimes raised me to ecstasy, but more frequently sunk me into the lowest dejection"[253]. These novels have a profound impact on the creature, stirring a range of emotions within it, from moments of ecstasy to periods of deep melancholy. This exposure to literature awakens in the monster a host of new images, feelings, and perspectives, expanding its intellectual and emotional horizons. However, the monster's relationship with knowledge is complex. While it finds these literary experiences often elevating, they also plunge it into deep lows. This highlights the multifaceted nature of knowledge, which can both uplift and burden the individual who possesses it. The monster's exclamation, "Of what a strange nature is knowledge!"[254] emphasises the ambivalence it feels toward the insights it gains through its encounters with literature and its own self-awareness.

This relationship between the monster and literature is something that is clearly seen in Yod's character. However, Yod begins on the path to self-agency soon after he is created, noting that is "mortal too"[255]. Yod's realisation of his own mortality marks a significant turning point. This understanding that he shares a fundamental aspect of existence with humans introduces the notion that he is not merely a machine, but a being with the capacity for profound experiences, emotions, and understanding. The contrast between Yod and the golem is highlighted through Dr. Yatsuko's assertion that Yod possesses "a mind but no consciousness"[256]. Yod's assertion that he has "as much consciousness as you do,"[257] challenges this perspective, signalling his

---

[251] Shelley 116
[252] Shelley 121
[253] Shelley 121
[254] Shelley 116
[255] Piercy *He, She and It* 94
[256] Piercy *He, She and It* 370
[257] Piercy *He, She and It* 370

growing awareness of his own unique identity. The dialogue between Yod and Dr. Yatsuko is pivotal, showcasing Yod's assertion of his own consciousness and agency. His claim that he has "as much consciousness"[258] as a human implies that he sees himself as possessing qualities that transcend the boundaries of mere programming. This sentiment foreshadows Yod's role in blurring the lines between human and machine, suggesting that he occupies a liminal space between the two.

Malkah states "Yod is a person. Persons cannot be sold. If you want him, you must hire him away of his own volition."[259], which reinforces his humanity, arguing that he should not be treated as property that can be bought and sold. This stance challenges the traditional view of machines as mere tools devoid of personal will. However, Dr. Vogt's counterargument reflects the scepticism prevalent in the story's world, as seen in the lines below;

> Machines do not have volition, Dr. Shipman—surely you have not entirely taken leave of your senses," Dr. Vogt said. "They have programming that defines goals. Since they are compelled to pursue those programmed ends, they may appear willful, but we are dealing with the same projection of affect my little boy was guilty of when he used to say a chair hit him.[260]

Dr Vogt states that machines, including Yod, lack true volition and are instead driven by programmed goals, comparing them to a child's imaginative projection. This reflects the tension between different perspectives on consciousness, free will, and the essence of personhood in the context of a society where advanced technology blurs these distinctions. However, it is Yod's extreme sense of human agency that allows him to have an intimate relationship with Shira. Their relationship first begins in the novel's cyberspace, known as the Base. The Base is known as "the gold mine of the town, where the systems were created that were the town's main export" [261], and Shira emphasises that "moving in and out of large AI databases was something taught all

---

[258] Piercy *He, She and It* 370
[259] Piercy *He, She and It* 370
[260] Piercy *He, She and It* 370
[261] Piercy *He, She and It* 46

children in Tikva"[262]. Shira's assertion that moving in and out of the Base is a common skill taught to all children in Tikva emphasises the level of technological fluency in their society. Additionally, it implies that their interactions in the virtual world are not just a novelty, but a fundamental part of their lives, reflecting the extent to which technology shapes their identities and relationships. Shira further states that interacting in the base is a product of imagination;

> In the image world, I am the power of my thought, of my capacity to create. There is no sex in the Base or the Net, but there is sexuality, there is joining, there is the play of minds like the play of dolphins in surf. [263]

Here, the parallel drawn between Shira and Yod's relationship in the Base and the concept of the Golem's creation is interesting. The "capacity to create" [264] through thought, as emphasised by Shira, is reminiscent of the ancient tale of the Golem, where Rabbi Loew's creation was shaped from inanimate matter to serve and protect. The power to shape and create through thought draws a subtle connection between the themes of technological agency in the cyberpunk world and the mysticism surrounding the Golem's story. Shira's description of the Base as a realm where she is the power of her own thoughts and the creator of her experiences captures the essence of the virtual space's significance. Within this space, the limitations of the physical world are transcended, and individuals have the agency to shape their experiences according to their imagination. Shira's emphasis on "joining" [265] and the "play of minds" [266] reflects the depth of the connections formed in this digital landscape, hinting at the emotional and intellectual bonds that can be forged in the virtual realm.

Shira describes her first intimate encounter with Yod while in the Base, and highlights that "in her mind was the idea that it was time to treat him as a person, fully, because he was nothing less" [267]. Yod's capacity to be seen as a human and not a machine is fully actualised at this moment. His compassionate and ethical personality allows Shira to form an intimate relationship

---

[262] Piercy *He, She and It* 141
[263] Piercy *He, She and It* 158
[264] Piercy *He, She and It* 158
[265] Piercy *He, She and It* 158
[266] Piercy *He, She and It* 158
[267] Piercy *He, She and It* 164

with him. However, it's important to acknowledge the multi-faceted nature of their relationship. Shira's use of Yod as a means to "fill all available mental space"[268] shows the intricate dynamics at play. While Yod's personhood and emotional qualities facilitate their connection, Shira's motivations also involve personal growth and self-improvement, possibly utilising Yod as a tool to better herself and her own experiences.

The analysis provided by Diane Sautter, in her article "Erotic and Existential Paradoxes of the Golem: Marge Piercey's "He, She and It"", offers further insight into Shira's mindset:

> At this point, Shira is treating Yod mechanistically: he is her instrument to a freer self, as if she were healing herself through making love with a machine in order to prepare for some future affair with a real man. In fact, it is Shira who has the most problem with stereotypes in the love-pairing, and she frequently thinks of Yod as a machine. She broods over his nature as a machine since, as their affair continues, others confront her with it, referring to Yod as a dildo. [269]

The idea that Shira treats Yod mechanistically to heal herself and prepare for future human relationships suggests a dual nature to their connection. On one hand, Shira acknowledges Yod's personhood and forms a genuine bond with him. On the other, she recognises his utility as a tool to navigate her own emotional landscape and potentially overcome personal obstacles. Sautter's observation about Shira's internal conflict regarding Yod's nature as a machine echoes the complexities inherent in forming relationships with non-human entities. This internal struggle reflects broader societal attitudes and biases, as Shira grapples with reconciling her genuine emotional connection with Yod against the backdrop of societal norms and perceptions.

In *He, She, and It*, Marge Piercy intricately weaves a narrative that not only challenges the traditional boundaries between humans and AI but also provides a profound commentary on the nature of empathy, autonomy, and identity. This chapter has explored the novel through a

---

[268] Piercy *He, She and It* 164

[269] Sautter, Diane. "Erotic and Existential Paradoxes of the Golem: Marge Piercey's "He, She and It."" *Journal of the Fantastic in the Arts*, vol. 7, no. 2, 1995, pp. 255-268. 259

cyberpunk lens, highlighting how Piercy juxtaposes the advanced technological world with deeply human experiences and emotions. The central relationship between Shira and Yod, a cyborg, stresses the complex dynamics of human-AI interactions, where lines of empathy, love, and moral consideration become blurred.

The novel's setting in a dystopian future offers a canvas where socio-political and ethical dilemmas of technology and its impact on humanity are vividly portrayed. In this regard, Piercy not only imagines a future where AI and cyborgs are integral to society but also prompts readers to consider the implications of such integration. The feminist perspective further enriches this discourse, as it foregrounds the role of women in shaping and interacting with technological and societal structures.

Through the story of Shira and Yod, along with the interwoven narrative of the Golem, Piercy delves into the existential paradoxes of artificial life. Yod's journey towards self-actualisation and his ultimate sacrifice highlights the capacity for AI to transcend their programmed nature, suggesting a form of emergent consciousness and empathy. Meanwhile, Shira's emotional journey with Yod and her ultimate decision not to recreate him is a poignant reflection on the ethics of creating sentient beings. In this chapter, the portrayal of AI characters as agents of their own narratives challenges traditional perceptions of AI as passive, emotionless tools. The chapter delves into how AI characters exhibit self-awareness, make independent decisions, and form complex emotional responses, thereby aligning with my thesis topic on AI's potential for empathy and ethical interactions. This depiction of AI agency not only humanises these characters but highlights questions about autonomy, consciousness, and the rights of artificial entities. By presenting AI characters as active participants rather than mere background elements, the chapter emphasises the need to reconsider our ethical and emotional engagement with AI in society.

He, She and It is a seminal work of feminist cyberpunk literature, offering a rich exploration of AI's role in society. Piercy's novel challenges readers to contemplate the evolving relationship between humans and technology, the nature of consciousness, and the moral responsibilities we hold towards beings we create. As AI continues to advance in the real world, the themes

explored in this chapter resonate more than ever, making *He, She, and It* a prescient and vital piece of literature in understanding our future with AI.

Chapter 4

"He struggled visibly to contain his emotion at seeing what he had wrought in pursuit of his father's vision: a child so wedded to machines that he could not acknowledge another human being." *- Dacey's Patent Automatic Nanny*

The Lifecycle of Software Objects:

The roots of science fiction as a literary genre are deeply intertwined with the works of authors like Jules Verne and H.G. Wells. Verne, with his imaginative foresight, explored space travel in *From the Earth to the Moon* (1865), captivating readers with the concept of journeying beyond our planet. This novel, alongside others by Verne, laid a foundation for the genre by combining scientific principles with adventurous storytelling. H.G. Wells further expanded the scope of science fiction with *The War of the Worlds* (1898), where he envisioned a chilling invasion of Earth by Martians. Wells' work often delved into the unknown and the fantastical, pushing the boundaries of what was considered possible at the time. His narratives, rich with speculative ideas, contributed significantly to the genre, often provoking thought about the potential and dangers of scientific advancements.

The term "science fiction" itself, however, only gained widespread use in the 20th century. This was largely due to the American magazine publisher Hugo Gernsback, who introduced the term "scientifiction" in the debut issue of Amazing Stories in 1926[270]. This term eventually evolved into what we now know as science fiction. Gernsback's vision was to create a genre that melded scientific facts with imaginative fiction, providing a new form of storytelling that was both entertaining and thought-provoking. This genre has since seen numerous interpretations and definitions, often described as an apparent oxymoron that blends the plausible with the fantastical, constantly pushing the boundaries of our imagination and scientific understanding.

[270]SF Dictionary. "scientifiction." *Historical Dictionary of Science Fiction*, 1 March 2021, https://sfdictionary.com/view/104/scientifiction. Accessed 4 January 2024.

If a story is grounded in science, one might ask, how can it be reconciled with narratives that are purely fantastic or imaginary? One answer comes from the American author Ray Bradbury, who argued that science fiction should concern itself with what is possible, rather than what is not. "I define science fiction as the art of the possible,"[271] he told the Christian Science Monitor in 1980, "Fantasy is the art of the impossible. Science fiction, again, is the history of ideas, and they're always ideas that work themselves out and become real and happen in the world." [272]

With this in mind, science fiction often parallels or predicts scientific developments that "become real and happen in the world", a prime contemporary example being the rise of AI chatbots that are capable of passing the Turing Test, which examines a machine's ability to exhibit intelligent behaviour indistinguishable from that of a human. The following excerpt comes from a conversation with a chatbot called PI, which is based on OpenAI's Large Language Model technology (Pi.ai):

> Can you pass the Turing Test?
>
> Haha! Love this question. I'm a chatbot designed to mimic human conversation, so I'm pretty sure I could pass the Turing Test if we were chatting about a broad range of topics. But here's the thing - I don't have a physical body, so if the test were to include physical tasks, I'd be SOL!

As the response shows, PI is capable of conducting conversations, in text or voice, that are rich in wit and sly self-awareness, if not sentience at this nascent phase of innovation. With these current modes of communication with AI, it becomes interesting to consider how literature foreshadows these events. One author who wrote a novel on this topic in 2019 was the American scientist fiction writer, Ted Chiang.

The fourth and final chapter of my dissertation will focus on two short stories by Ted Chiang, "The Lifecycle of Software Objects", and "Dacey's Patent Automatic Nanny". "The Lifecycle of Software Objects", Chiang's novella, centres around the idea of keeping AI robots as "pets". In the story, the protagonist, Ana, is a zookeeper who is head-hunted to train digital entities, known

---

[271]Unger, Arthur. "Ray Bradbury: the science of science fiction - CSMonitor.com." *The Christian Science Monitor*, 13 November 1980, https://www.csmonitor.com/1980/1113/111356.html. Accessed 4 January 2024.
[272] Brandbury qtd. in Unger 1980

as digients, to "pitch them as pets you can talk to, teach to do really cool tricks." [273]. This fictional portrayal of virtual pets finds an analogy in video games such as Sims, which allows the player to control the lives of virtual humans and their animal companions. There are similarities between the digients and Tamagotchi, a handheld device which rose to popularity in the early 2000s, allowing users to "hatch" a pet on screen, and take care of it as it ages. Ana adopts one of the Blue Gamma digients, Jax, which first appears as an avatar made to represent a "neo-Victorian robot made out of polished copper" [274].

In this near-future world, the concept of digients introduces a unique twist on AI development. These digital creatures start as simple beings but can evolve, learn, and develop complex personalities over time. The story follows two protagonists, Ana and Derek, who become involved in the creation and care of these digients. Ana, in particular, forms a strong emotional bond with her first digient, Jax. This relationship serves as a central theme throughout the narrative. Ana works on training the digients to develop the best traits for human interaction. One of these examples involves removing the digients' jealousy trait,

> The digients love attention, and if one notices that Ana's giving praise to another, it tries to get in on the action. Most of the time this is fine, but whenever a digient seemed particularly resentful of its peers or of Ana, she would flag it, and its specific genome would be excluded from the next generation.

Ana likens this process to "breeding dogs", as well as "baking endless batches of brownies and sampling each one's toothsomeness to find the perfect recipe". Just as a baker tirelessly experiments with ingredients and methods to achieve the ideal brownie, Ana engages in a relentless pursuit of crafting the perfect digient, one that embodies the most desirable human traits for interaction. This comparison draws attention to the experimental nature of her endeavours, where each generation of digients is a step closer to achieving her ideal vision of digital companionship. Ana's analogy of "breeding dogs" emphasises the deliberate and selective

---

[273] Chiang, Ted. "The Lifecycle of Software Objects". Chiang, Ted. *Exhalation*. New York: Penguin Random House, 2019. 62-163. Pg 65

[274] Chiang, "Lifecycle of Software Objects" 71

process of refining the digients' traits, much like a breeder striving to produce specific qualities in a canine lineage. This comparison emphasises the degree of precision and intentionality involved in crafting these digital creatures. Their development is not left to chance but is carefully guided by human intervention.

As the story unfolds, readers witness the challenges Ana and Derek face in raising and training their digients. They devote significant time and effort to teaching their digital creations, and in return, the digients develop unique identities and exhibit signs of consciousness and sentience. This human-digient relationship blurs the lines between what is real and what is artificial, as well as challenging preconceptions about AI and its potential to form meaningful connections. However, the novella takes a darker turn when the technology and platforms that support digients become obsolete. Ana and Derek are confronted with the prospect of their beloved digients facing abandonment or even digital death. This situation forces them to confront profound ethical questions: What are the rights of these digital beings? What responsibilities do the creators bear towards their creations? Can sentient digital entities be considered as deserving of the same ethical considerations as living beings?

The novelty of the digients, similar to that of a Tamagotchi, soon wore off, and "caused general interest in digital life-forms to dwindle" [275]. In his article, "Book and Media Recommendations: What If?"; The Lifecycle of Software Objects; Sapiens: A Brief History of Humankind; and Gulp: Adventures on the Alimentary Canal", Brittland DeKorver discusses how "Ana and other stalwart digient owners faced crippling shifts in technology and a society that could not comprehend their devotion to their digient companions" [276]. One of the most important themes in this story, as highlighted by critic Sunyoung Ahn in "The Everyday Life of Artificial Intelligence: The Humanism of Ted Chiang's 'The Lifecycle of Software Objects'", is that of "empathic imagination" [277]. The ability to be consciously aware of the feelings of the digients is woven throughout the story. This chapter will aim to discuss how Ana's nurturing position over

[275] Chiang, "Lifecycle of Software Objects" 104
[276] DeKorver, Brittland. "Book and Media Recommendations: What If?; The Lifecycle of Software Objects; Sapiens: A Brief History of Humankind; and Gulp: Adventures on the Alimentary Canal." *Journal of Chemical Education* (2015): 1143-1145. Pg 1144
[277] Ahn, Sunyoung. "The Everyday Life of Artificial Intelligence: The Humanism of Ted Chiang's The Lifecycle of Software Objects." *Science Fiction Studies* 47 (2020): 73-92. Pg 74

the digients allows the reader to bond emotionally with these nonhuman machines. After analysing the story, I will turn to focus on the ethical questions which arise from it. Such questions will follow in similar lines: 1) What moral obligations, if any, do we have in a human-robotic relationship? 2) If a moral obligation is required, will this be in the interest of the humans, or the robots?

The concept of virtual pets, prominently depicted in "The Lifecycle of Software Objects," provides a fascinating lens to examine our relationship with artificial entities. These digital companions, like the "digients," offer unique insights into how we perceive and connect with technology. The Tamagotchi, one of the earliest representations of this phenomenon, captivated users with its demands for care and attention, highlighting the human desire to nurture even non-traditional companions. Hidden in my drawer is a long-forgotten Tamagotchi, a gift from my younger years. My short-lived obsession with it echoes the global trend; despite its meteoric rise in popularity, the Tamagotchi soon faced a decline. According to Bloch (1999), demand for the toy plummeted, with prices dropping below cost. Such digital entities, while not tangible like traditional pets, emphasise our inclination towards objects that require our care. Jean-Loup Rault, in his article "Pets in the digital age: live, robot, or virtual?," states that while robotic pets can elicit human responses similar to live ones, their virtual lives are fragile, especially when neglected [278].

By the 21st century, virtual pets, particularly the Tamagotchi, showcased the potential of forming emotional connections with digital entities. This concept was further explored in Chiang's "The Lifecycle of Software Objects," where "digients", powered by AI, act as digital companions capable of evoking empathy. Such virtual entities not only intrigue due to their novelty but also prompt questions about our evolving relationship with technology, the real-digital boundary, and the ethical implications of these interactions. The progression from early virtual pets to sophisticated digients in literature indicates the expanding scope of compassionate AI. Authors like Ted Chiang delve into human emotions and the depth of our connections with artificial entities. Yet, the fleeting nature of such fascinations, as seen with the Tamagotchi's declining

[278] Rault, Jean-Loup. "Pets in the digital age: live, robot, or virtual?" *Frontiers in Veterinary Science*, vol. 2, no. 11, 2015, pp. 1-3, doi: 10.3389/fvets.2015.00011. 1

popularity, mirrors the narrative in Chiang's work. The initial allure of digital beings often gives way to broader reflections on our complex and evolving relationship with technology. In his book *Discognition*, Steven Shaviro argues that our understanding of consciousness, particularly when it comes to non-human entities such as AI, is largely limited to our own human experiences and understandings[279]. Shaviro states that digients are "recognizable descendants of such currently-existing "software objects" as Tamagotchis, chatbots, non-player characters in video games, and "digital assistants" like Siri"[280]. However, the difference between these two are what Shaviro describes as "a matter of generality." [281]. He states that "Chiang's digients have a well-rounded overall sensibility, rather than any particular skills. In contrast, digital agents today are limited to certain specialized tasks, like recognizing faces, translating text, driving cars, and playing games like chess and Jeopardy"[282].

The fading allure of the Tamagotchi compared to the digients offers an interesting exploration in the evolution of digital companionship. The Tamagotchis, once a craze, became easy to disregard when their novelty wore off. Their simplicity allowed for a clear-cut detachment; when their digital lives ceased, so too did our responsibility. They were pocket-sized distractions, not entities with whom one could have a profound emotional connection. On the other hand, digients, despite facing a similar decline in popularity, evoke a more complex response. These advanced AI entities were designed to be more engaging, more demanding, and, as a result, formed a stronger bond with their human counterparts. Customers deemed them "too demanding" [283], a testament to their complex programming that simulated needs similar to living beings. The waning interest in digients was not merely a trend passing but a phenomenon with emotional ramifications. Even as the majority of owners found the responsibilities overwhelming and left their digients in a state of digital limbo, a dedicated fraction of owners persisted. This lingering attachment suggests that the digients managed to cross a psychological threshold, transforming from mere digital entities to companions worthy of enduring care. The survival of a "small community of hardcore digient owners"[284] amidst Blue Gamma's economic failure

---

[279] Shaviro, Steven. *Discognition*. London, Repeater, 2016. 15
[280] Shaviro 60
[281] Shaviro 60
[282] Shaviro 60
[283] Chiang "Lifecycle of Software Objects" 85
[284] Chiang "Lifecycle of Software Objects" 89

emphasises a deeper human-digital connection. Unlike the Tamagotchis, digients were not so easily abandoned, signifying a shift in how these virtual entities were perceived and valued. For Ana, the closure of Blue Gamma is more than an end of service—it resonates with the memory of a zoo's closure, where the lives of real animals were affected. Her comparison reveals a level of empathy for the digients that transcends their virtual existence, viewing them through a lens of sentient worth rather than as inanimate tech products.

As the short story progresses, the reader is introduced to Real Space, a virtual world hailed as "the hotspot of digital terrain formation"[285]. This new realm symbolises the advancing frontier of virtual environments, providing a sophisticated backdrop where the digital and the real continue to blur and merge. Accompanying the rise of Real Space is the development of a groundbreaking software named Sophonce, which marks a pivotal evolution in the interaction between humans and digients. Sophonce represents a departure from the need for direct human interaction, as it enables digients to be "taught via software"[286]. This innovation presents both an opportunity and a dilemma. The digients, previously reliant on human engagement to develop their personalities, now risk becoming "charmless" [287] due to the automated nature of their education. The software's focus is on efficiency, grooming digients to become "useful workers within a matter of months"[288]. As Shavrio notes, "The "lifecycle" of the story's title is not only that of the digients themselves, but also of the corporations that build and develop them and market them. In a very real sense this "lifecycle" is a commercial product cycle. Despite their sentience, the digients are threatened with obsolescence like any other bit of software"[289]. This utilitarian approach prioritises practical skills over the unique personalities that emerge from organic interactions, potentially reducing digients to mere tools rather than autonomous entities with individual characteristics.

Within this evolving digital landscape, Ana's perspective remains firm. While she sees the practicality in allowing digients to acquire new skills, she is adamantly against commodifying

---

[285] Chiang "Lifecycle of Software Objects" 104
[286] Chiang "Lifecycle of Software Objects" 105
[287] Chiang "Lifecycle of Software Objects" 105
[288] Chiang "Lifecycle of Software Objects" 105
[289] Shaviro 74

them. Her stance, "maybe Jax can make money, but Jax isn't *for* making money" [290], speaks to a deeper conviction that digients, much like the creatures that once inhabited the zoo, should not be exploited for financial gain. For Ana, the intrinsic value of digients extends beyond their utility. Her resistance to selling Jax, despite the digient's potential profitability, is a stand against the exploitation of beings she views as more than just sophisticated programs; to her, they are entities worthy of dignity and respect. Integrating this narrative into the broader context of digital companionship enriches the discussion. The emergence of Real Space and Sophonce highlights the relentless march of technological innovation, while Ana's reaction to these developments emphasises the ethical quandaries they present. This contrast between the march towards efficient, utilitarian applications for digients and the heartfelt resistance to their commercial exploitation emphasises the tension between technological advancement and the preservation of empathetic, meaningful connections.

Therefore, the waning interest in digients, juxtaposed against the relentless innovation represented by Real Space and Sophonce, paints a complex picture of our future with artificial intelligence. It asks us to consider how we might balance the practical applications of such technologies with the compassionate ethos that characters like Ana embody. Just as the fleeting nature of Tamagotchis' popularity points to a throwaway culture in the digital age, the reaction to the digients' plight reflects a deepening struggle over the soul of our interactions with increasingly sophisticated artificial entities. Ana's empathy, reminiscent of the sorrow felt over the closure of a zoo, positions her as a guardian of a fading era of digital innocence, where each entity is seen as having intrinsic worth beyond their utility or novelty.

In part seven of the narrative, the passage of time brings with it the harsh realities of technological obsolescence. Two years and two months later, the digients find themselves at a crossroads, their existence threatened by incompatibility with the updated Real Space. Ana, alongside fellow digient caretakers, is marooned in a relic of digital space—Data Earth's private server—a refuge that comes at a cost. The financial burden of maintaining this sanctuary looms large, with failed fundraising efforts casting a shadow over the digients' future. The reader is then introduced to an upcoming company called Polytope, "which has announced plans for

[290] Chiang "Lifecycle of Software Objects" 109

launching an enormous breeding program to create the next Andro" [291]. Polytope offers Ana and the other digient owners a deal, "in exchange for paying to port the Neuroblast engine, Polytope would get a percentage of any income generated by the digients in perpetuity"[292]. Here, the group becomes hopeful that the digients will have a future in real space, however "the company's answer was no; the only digients that Polytope is interested in are Sophonce digients, whose obsessive focus is a necessity if they're going to replace conventional software." [293]. Some owners contemplate a temporary suspension of the digients, a digital hibernation, until the funds can be sourced for their transition—a port to the new version of Real Space.

Ana's response to this proposition is both poignant and telling. She resists the notion of suspension, fearing it could become a permanent solution masquerading as a temporary fix. Her words, "It's just too easy to get into perpetual postponement mode"[294], emphasise her commitment to the digients. For Ana, keeping Jax conscious—awake in the digital realm—is a source of purpose, a catalyst driving her fundraising endeavours. Jax, a creation of ones and zeroes, becomes an anchor for her motivations, a being whose virtual heartbeat synchronizes with the rhythms of her daily life. Ana's compassion stands in stark contrast to the broader societal trend of "suspending" digients without a second thought. The digients' plight is a mirror to the disposable nature of technology and the ethical considerations it raises. Can a digital entity, endowed with the illusion of life, be ethically "suspended", or does their apparent sentience demand more from us? This anthropomorphism, the attribution of human traits to non-human entities, blurs the lines of moral responsibility. While the world may have moved on from the digients, Ana, and those like her, represent a bridge between human empathy and digital existence, challenging the transient nature of our interactions with technology and provoking a conversation about the sustainability of digital companionship.

Jax's own awareness of the financial strain introduces a turn in the story. His wish to be suspended - to be spared the waiting and reawaken when the time is right - mirrors the all-too-human desire to escape hardship. This is seen as he states, "Sometimes wish I just be suspended,

---

[291] Chiang, "Lifecycle of Software Objects" 128
[292] Chiang, "Lifecycle of Software Objects" 129
[293] Chiang, "Lifecycle of Software Objects" 129
[294] Chiang, "Lifecycle of Software Objects" 129

not have to wait more. Restarted when I can enter Real Space, feel like no time passed"[295]. That Jax articulates such a thought independently, without influence from the user-group forums, is a testament to his advanced cognitive development. This moment is a reflection of the emergent human-like consciousness within these digital beings, emphasising the theme of sentience in the novel.

Ana's response to Jax highlights the crux of their relationship, as she states, "My life might be simpler if I didn't have you to take care of, but it wouldn't be as happy. I love you, Jax."[296]. These words encapsulate the essence of compassion extending beyond the boundaries of flesh and blood, reaching into the realm of artificial intelligence. Ana sees Jax not merely as a collection of algorithms or a sophisticated program, but as a genuine presence in her life, capable of reciprocating feelings and deserving of love. This exchange is a cornerstone of the story, one that vividly paints the emotional bond between humans and AI. It brings into sharp focus the ethics of digital companionship, the responsibilities of creators towards their creations, and the attachment that can form even with beings made of code.

The narrative takes a significant turn when Ana is presented with a job opportunity at Polytope. The offer to lead their team, "training Sophonce digients"[297], comes with the kind of financial security that is hard to come by: "great salary, three years guaranteed employment, and a signing bonus"[298]. Yet, this seemingly golden opportunity is tarnished by a morally dubious requirement—the use of InstantRapport. "InstantRapport,", as Ana explains, "is one of the smart transdermals, a patch that delivers doses of an oxytocin-opioid cocktail whenever the wearer is in the presence of a specific person. It's used to strengthen rocky marriages and strained parent-child relationships, and it's recently become available without a prescription"[299]. The fact that such a tool is needed to foster affection between trainers and Sophonce digients raises ethical questions about the authenticity of the resulting relationships. The company's rationale is clear:

---

[295] Chiang "Lifecycle of Software Objects" 130
[296] Chiang "Lifecycle of Software Objects" 130
[297] Chiang "Lifecycle of Software Objects" 156
[298] Chiang "Lifecycle of Software Objects" 131
[299] Chiang "Lifecycle of Software Objects" 131

they believe that trainers will achieve better results if they have genuine affection for the digients, even if that affection is pharmacologically induced.

Derek's astonishment upon learning of this requirement reflects the unsettling nature of using drugs to manipulate emotions in a professional setting. InstantRapport was initially intended for personal relationships, such as mending familial bonds or marital troubles, not for enforcing workplace performance. This blurring of lines between professional ethics and personal life is emblematic of the broader issues that arise when advanced AI is involved. Ana's conundrum at Polytope is a microcosm of this larger debate. It forces the reader to decide whether the emotional connections with AI, chemically enhanced or not, are valid. Furthermore, it compels us to confront the reality that our technological advancements might be outpacing our ethical frameworks.

The job offer's ethical complexity deepens when Ana is reminded of the persistent interest in using digients for adult entertainment. Jennifer Chase's approach, representing Binary Desire, echoes past inquiries about sexual uses for the digients. This is not an isolated phenomenon; as it is noted that "as long as there have been digients, there have been people trying to have sex with them"[300]. This mirrors broader themes in science fiction where AI entities are not only companions but also objects of desire, as seen in characters like Adam from Ian McEwan's *Machines Like Me*, designed to fulfill mature roles for their owners, or Yod from Marge Piercy's *He, She and It*, capable of intimate human relationships.

As Ana exhausts every avenue to integrate the digients into Real Space, she approaches "Exponential Appliances"[301], envisioning a new role for the digients as "a butler that is utterly loyal and attentive from the moment it's switched on"[302]. However, the company's rejection reflects a broader theme in the story: the struggle to define the essence of digients in a world fixated on clear-cut roles and utilitarian purposes. The owner's statement, "we aren't looking for superintelligent employees, we're looking for superintelligent products"[303], encapsulates a

[300] Chiang "Lifecycle of Software Objects" 133
[301] Chiang "Lifecycle of Software Objects" 133
[302] Chiang "Lifecycle of Software Objects" 133
[303] Chiang "Lifecycle of Software Objects" 153

prevalent view in the business world: the preference for products over entities that evoke a sense of personhood. For them, the digients, imbued with years of learning and personality development, are too individualistic, too personable, to fit into their product line.

This conundrum leaves the digients in a limbo, perceived as "not robotic enough" for Binary Desire's purposes and "too sentient" for Exponential Appliances. It is Derek's affection for Ana that brings a semblance of resolution. Derek recognises Ana's moral dilemma with the Polytope job offer and its requirement to use InstantRapport. With a difficult moral decision, he opts to offer Marco to Binary Desire instead, wrestling with the ethical question: is it "better for Marco to have his brain chemistry altered than for Ana to alter hers?"[304]. While it is clear that Ana would not agree to this, Marco states that "Agree wrong if they edit digients not want be edited. But if digient agree before be edited, then not wrong" [305], thus giving his consent to being part of Binary Desire. The story ends with Ana feeling hurt at Derek's decision, however she attempts to see the optimistic side of things, as seen in the lines below;

> She imagines Jax maturing over the years, both in Real Space and in the real world. Imagines him incorporated, a legal person, employed and earning a living. Imagines him as a participant in the digient subculture, a community with enough money and skills to port itself to new platforms when the need arises. Imagines him accepted by a generation of humans who have grown up with digients and view them as potential relationship partners in a way that members of her generation will never be able to. Imagines him loving and being loved, arguing and compromising. Imagines him making sacrifices, some hard and some made easy because they're for a person he truly cares about. [306]

In these closing lines, Ana's imagination takes flight, envisioning a world where Jax and other digients not only survive but thrive. She sees Jax evolving, growing in both the virtual realm and the real world. Her vision extends to him achieving legal personhood, gaining employment, and becoming an integral part of a self-sustaining digient community. Moreover, she imagines a

---

[304] Chiang "Lifecycle of Software Objects" 158
[305] Chiang "Lifecycle of Software Objects" 147
[306] Chiang "Lifecycle of Software Objects" 162

societal shift where future generations view digients not just as advanced technology but as potential partners in life, capable of love, arguments, compromises, and sacrifices. This vision is a testament to the depth of Ana's empathy and her belief in the inherent value of the digients as more than just digital creations.

Ana's dream for the digients encapsulates the core themes of the story: the evolving relationship between humans and AI, the struggle to balance technological advancement with ethical considerations, and the enduring quest for understanding and acceptance in an increasingly digital world. It highlights the potential for AI to transcend its programmed boundaries, suggesting a future where digital and human lives are intertwined in complex, meaningful ways. The story leaves the reader questioning the intricate moral and emotional landscapes we navigate as we forge deeper connections with the artificial beings we create, challenging us to redefine our understanding of sentience, companionship, and the very nature of existence in an age where the lines between the real and the digital continue to blur.

"The Lifecycle of Software Objects" highlights the psychological and emotional complexities of AI, particularly through the character of Ana and her interactions with digients. These AI entities are not just sophisticated programs; they exhibit a form of emotional depth that challenges conventional views of AI as emotionless. The story aligns with my thesis by presenting a scenario where AI can form meaningful relationships and demonstrate empathy, highlighting the potential for ethical considerations in AI development and interaction. This narrative contributes significantly to the discussion about the evolving nature of AI and its capacity for emotional and ethical engagement.

The narrative forces the reader to consider the weight of such connections and the moral obligations they entail. As the characters grapple with the practicalities of sustaining the digients, the story urges us to contemplate the intangible yet invaluable aspects of joy and love that these entities bring into human lives. It is a dance of dependency and affection, a testament to the depths of human empathy, and perhaps, a mirror showing us the lengths to which we will go to preserve the things we care for, irrespective of their origin in the physical or digital world.

Dacey's Patent Automatic Nanny

The next story to be explored in this chapter is "Dacey's Patent Automatic Nanny". This short story follows Reginald Dacey's creation of the automatic nanny to raise children. The advertisement posted in a newspaper for the nanny reads as follows;

> "Do not leave your child in the care of a woman whose character you know nothing about. Embrace the modern practice of scientific child-rearing by purchasing DACEY'S PATENT AUTOMATIC NANNY"[307]

The use of AI to create a robot companion for children is a major theme in this story, which I have discussed in my analysis of *Klara and the Sun*. However, in this story, as opposed to Klara, Dacey is forced to prove his theory that robots can adequately raise children better than humans after one of the robot nanny's malfunctions and "routinely beat[s] the boy and administered Gregory's Powder (a potent and vile-tasting laxative) as punishment" [308]. In order to stand by his invention, Dacey adopts a child and raises it only using the robot nanny, resulting in a child who "was unresponsive to his commands, and shortly afterwards a physician diagnosed the child as "feebleminded."" [309]. The child, Edward, became responsive only to machines, and not to humans. "Dacey's Patent Automatic Nanny" represents an extreme in child-rearing with machines, focusing solely on responsiveness to technological care, which contrasts with the approach presented in *Klara and The Sun*, where the interaction between children and machines is more nuanced and human-like. Josie experiences no difficulties in assimilating her AF into the real world, while Edward is unable to communicate with humans until the nurses used "an intercom system that emulated the low-fidelity audio of the original Automatic Nanny's gramophone" [310]. Although this narrative does not include a clear example of constructing empathy in AI, I have included it in my dissertation due to its potential for a parallel analysis

---

[307] Chiang, Ted. "Dacey's Patent Automatic Nanny." Chiang, Ted. *Exhalation*. New York: Penguin Random House, 2019. 164-174. Pg 167
[308] Chiang "Dacey's Patent Automatic Nanny" 174
[309] Chiang "Dacey's Patent Automatic Nanny" 180
[310] Chiang "Dacey's Patent Automatic Nanny" 173

with *Klara and The Sun*. I aim to discuss the effects of nurturing children with AI, both positive, as we see with Klara, and negative, as we see with Dacey's robotic nanny.

In 1955, Philip K. Dick released a short story titled "Nanny". The story is set in an undefined future, where every family owns a mechanical nanny. The nanny performs every household duty for the Fields family, ensuring that Mrs Fields "did not have to wake the children up, dress them, see that they were washed, ate their meals, or anything else. She did not even have to take them to school"[311]. Nanny's role transcends mere functionality; it extends into the realm of companionship and affection. The narrative underscores this when Philip K. Dick writes, "Both children loved her,"[312] providing a testament to the genuine bond that has formed between the mechanical nanny and the children. This affection is further emphasised when Nanny requires a trip to the repair shop, causing the children to "cr[y] and cr[y] without stopping"[313]. Physically, the nanny is described in the lines below;

> Nanny was built in the shape of a sphere, a large metal sphere, flattened on the bottom. Her surface had been sprayed with a dull green enamel, which had become chipped and gouged through wear. There was not much visible in addition to the eye stalks. The treads could not be seen. On each side of the hull was the outline of a door. From these the magnetic grapples came, when they were needed. The front of the hull came to a point, and there the metal was reinforced. The extra plates welded both fore and aft made her look almost like a weapon of war. A tank of some land. Or a ship, a rounded metal ship that had come up on land. Or like an insect. A sowbug, as they are called.[314]

One of the most striking aspects of Nanny's design is the pointed front of its hull, which is reinforced with additional metal plates. It resembles a "weapon of war" [315], suggesting the possibility of a tank or a ship that has somehow ventured ashore. This comparison of the machine to a weapon further foreshadows the violence that the Nanny hides. At night, Nanny

---

311 Dick, Philip K. *Nanny*. Citadel Twilight, 1955. 3
312 Dick 3
313 Dick 3
314 Dick 5
315 Dick 5

would escape the Field's home and engage in fights with neighbouring mechanical nannies, describing it as "The wild frenzy of battle" [316]. However, at the same time, the machine is compared to an insect, or a "sowbug"[317]. This varying comparison offers a juxtaposition in the way AI is viewed - both as a violent weapon and as a harmless insect. The spherical shape of the robot is interesting as it emphasises the unconventional design of humanoids. In contrast to Nanny, later depictions of mechanical caregivers, such as the Automatic Nanny in Chiang's "Dacey's Patent Automatic Nanny," embrace a more human-like appearance. Here, the nanny is described as more humanlike, with "arms [that] formed a cradle for rocking the baby" [318]. However, Dick's narrative of creating a domestic assistant is furthered in Chiang's story, as the nanny is advertised to teach the baby a "precise schedule of feeding and sleeping"[319], as well as "sooth[ing]"[320], having the ability to "work night and day"[321], and not "expos[ing] your child to disreputable influences"[322].

When examining the two narratives side by side, an interesting progression becomes evident in the portrayal of mechanical caregivers—from the unconventional spherical design of Nanny to the more anthropomorphic and comprehensive models depicted in Chiang's story. Both stories, however, delve into the complex interplay between technology and humanity, exploring the profound impact of these mechanical nannies on family dynamics and the nurturing of children in futuristic worlds.

In "Dacey's Patent Automatic Nanny", a mathematician named Reginald Dacey creates the robotic nanny after discovering that the care from a human is not sufficient for his son. Reginald believes that parents should have little engagement in their children's life, stating that "Children are not born sinful, but become so because of the influence of those whose care we have placed them in" [323]. Thus, a robotic nanny should be able to care for a child neutrally, without adding

---

[316] Dick 9
[317] Dick 3
[318] Chiang "Dacey's Patent Automatic Nanny" 166
[319] Chiang "Dacey's Patent Automatic Nanny" 167
[320] Chiang "Dacey's Patent Automatic Nanny" 167
[321] Chiang "Dacey's Patent Automatic Nanny" 167
[322] Chiang "Dacey's Patent Automatic Nanny" 167
[323] Chiang "Dacey's Patent Automatic Nanny" 166

any influence in their life. Shortly after the release of Reginald's Automatic Nanny, the machine lost popularity after "an infant named Nigel Hawthrone was fatally thrown from an automatic Nanny when its mainspring snapped" [324]. Despite his efforts, Reginald was unable to gain the public's trust again, and retired back to his career in Mathematics. Years later, Reginald's son Lionel discovered a newspaper article describing his father's invention as "a monstrous contraption whose inventor surely despised children" [325]. In order to clear his father's name, Lionel set out an experiment of raising a child using only the Automatic Nanny.

Edmund, Lionel's illegitimate child, was "successfully raised by the Automatic Nanny until [he] was two years old, the age at which Lional Dacey felt it appropriate to switch him to human care" [326]. After being unable to care for him, Lionel admits Edmund to the "Brighton institute of Mental Subnormality"[327]. Here, it is discovered that Edmund was not developing as a normal child, as "although he was five, his height and weight were those of the average three-year-old"[328]. Edmund was then diagnosed with "psychological dwarfism"[329], and it was found that what he "needed was not more contact with a person, but more contact with a machine"[330]. The doctors at the institute adapted to care for Edmund by using "mechanical arms"[331], and a "low-fidelity audio of the original Automatic Nanny's gramophone"[332]. By disguising themselves as machines, the nurses were able to properly care for Edmund. This shows an unconventional way of nurturing in the human-robotic relationship, as it counters the idea that machines inherently lack the capacity for genuine care and connection with humans. Lionel's experiment with Edmund challenges the prevailing belief that solely human interaction is vital for a child's development.

It is interesting to note the comparison between the Automatic Nanny, and Klara from Ishiguro's *Klara and the Sun.* In "Dacey's Patent Automatic Nanny," the robotic caregiver is explicitly

[324] Chiang "Dacey's Patent Automatic Nanny" 169
[325] Chiang "Dacey's Patent Automatic Nanny" 169
[326] Chiang "Dacey's Patent Automatic Nanny" 171
[327] Chiang "Dacey's Patent Automatic Nanny" 170
[328] Chiang "Dacey's Patent Automatic Nanny" 171
[329] Chiang "Dacey's Patent Automatic Nanny" 171
[330] Chiang "Dacey's Patent Automatic Nanny" 173
[331] Chiang "Dacey's Patent Automatic Nanny" 173
[332] Chiang "Dacey's Patent Automatic Nanny" 173

designed to assume the role of a nanny, with a primary focus on childcare and domestic tasks. Reginald Dacey's creation is intended to provide infants with comprehensive care, including feeding, soothing, and constant supervision. However, as the story unfolds, it becomes apparent that the Automatic Nanny's rigid and mechanical approach to childcare has limitations, especially in terms of emotional and psychological development, as seen in Edmund's case.

On the other hand, Klara represents a more contemporary and nuanced exploration of AI companionship. While not explicitly designed as a nanny, Klara is a highly advanced AI who takes on various domestic duties and companionship roles within the household. Klara's relationship with Josie, the child she cares for, evolves over time as she learns and adapts to the family's dynamics. One notable distinction between the two AI caregivers lies in their level of autonomy and adaptability. Klara, equipped with advanced learning capabilities and sensory perception, is better positioned to adapt to the changing needs of her human companion. She grows alongside Josie, developing a deep emotional bond and understanding of human emotions. Unlike Edmund, Josie does not display cognitive or physical development problems as she matures alongside Klara, suggesting that Klara's capacity for empathetic and flexible caregiving contributes positively to Josie's development.

The comparison highlights the shift in the portrayal of AI caregivers in literature. While the Automatic Nanny in Chiang's story represents an earlier vision of rigid and limited AI caregiving, Klara embodies a more contemporary and sophisticated approach to AI companionship that prioritises emotional and psychological support. Both stories offer valuable reflections on the evolving dynamics between humans and AI, raising questions about the potential and limitations of machines in nurturing and caring for individuals in unique and complex ways. In weaving together Ted Chiang's "The Lifecycle of Software Objects" and "Dacey's Patent Automatic Nanny", there is an exploration of the evolving relationship between humans and artificial intelligence. Chiang's narrative delves into the complex emotional bonds formed between humans and their digital pets, reflecting on the depth of care and attachment that can develop in virtual spaces. This theme resonates intriguingly with the story of "Dacey's Patent Automatic Nanny", where AI takes a more physically tangible form as a robot caregiver. Both stories address the ethical, emotional, and psychological implications of entrusting AI with

roles traditionally held by humans, revealing both the potential benefits and pitfalls of such relationships. The contrast between Chiang's nuanced, often positive portrayal of AI-human interaction and the more cautionary tale in "Dacey's Patent Automatic Nanny" offers a multifaceted view of our future with artificial intelligence as both nurturer and companion.

The narratives in "Dacey's Patent Automatic Nanny", Philip K. Dick's "Nanny", and Kazuo Ishiguro's *Klara and the Sun* delve into the role of AI as caregivers, presenting a significant theme in literature. Each story uniquely examines the dynamics of artificial intelligence in nurturing roles, highlighting the complex interactions between humans and AI. This theme highlights the evolving understanding of AI's place in human life, especially in contexts traditionally reserved for human empathy and care. "Dacey's Patent Automatic Nanny" serves as a cautionary tale about the emotional and developmental risks of over-relying on rigid, unempathetic AI for childcare. In contrast, Dick's "Nanny", despite its more affectionate portrayal of AI, subtly hints at the underlying dangers and complexities of such relationships. On the other hand, *Klara and the Sun* offers a more optimistic view, showcasing the potential of AI to adapt, learn, and form genuine emotional bonds with humans. Together, these narratives reflect our evolving understanding of AI and challenge us to consider the ethical, emotional, and developmental implications of integrating artificial intelligence into the most intimate aspects of human life.

Conclusion

This dissertation set out to explore and discuss the relationship between simulated empathy and artificial intelligence as depicted in four literary works in the literary science fiction genre, as well as the ethical implications in these relationships. The works are: *Klara and the Sun* by Kazuo Ishiguro, *Machines like Me* by Ian McEwan, *He, She and It* by Marge Piercy, and two short stories by Ted Chiang. Each of these texts presents characters and situations that cast AI in literature in a new light, focussed on empathy and compassion. This perspective on AI is contrasted with the traditional portrayals found in classic science fiction. By contrasting these contemporary narratives with foundational works such as Isaac Asimov's *I, Robot,* Philip K. Dick's *The Mechanical Nanny*, and Mary Shelley's *Frankenstein*, the research portrays a clear evolution in the literary depiction of AI. Where once AI was seen primarily through a lens of utilitarian function or ethical dilemma, modern narratives have shifted towards a more nuanced representation, highlighting AI's potential for emotional depth and empathetic understanding.

This shift is not merely a change in storytelling, as it reflects a broader transformation in our societal and philosophical understanding of AI. Through these literary explorations, this dissertation examines how contemporary authors envision AI as autonomous agents and as entities capable of complex emotional experiences and moral reasoning. This change in perspective is indicative of a larger conversation in literature and technology about the nature of consciousness, the capacity for empathy in non-human entities, and the evolving relationship between humans and the increasingly sophisticated machines they create. My argument aims to serve as a bridge connecting classical and modern narratives of AI, offering insights into how our perception of artificial beings has matured and expanded over time. It invites readers to contemplate the expanding boundaries of empathy and the potential for AI not only to mimic human emotions, but to genuinely participate in the emotional foundation of our lives.

In Chapter 1, I followed the journey of Klara, the Artificial Friend to 11-year-old Josie, in Ishiguro's *Klara and the Sun*. Here, my main analysis consists of the idea that Klara, despite her naivety, is capable of displaying care and compassion towards Josie. Klara goes to extreme lengths to care for Josie during her health crisis. However, this chapter further emphasises the

ethical dilemma surrounding the replacement of human care and companionship with AI. Klara illustrates that while AI entities can be programmed to provide care and emotional support, they lack the capability to prioritise their own well-being, which presents a significant moral challenge. This challenges the notion that there is nothing unique about humans that AI cannot replicate. Klara's experiences demonstrate that although AI can imitate human emotions and caregiving, they do not possess the self-awareness and self-preservation instincts inherent to humans. This leads to a situation where AI could be exploited or discarded without regard for their emotional or "mental" state.

Further, Ishiguro presents the theme of empathy in the novel as a two-way street between humans and AI. Klara shows empathy towards humans, challenging our conventional understanding of AI as emotionless machines. Her interactions with such characters as Josie and Rick reveal a capacity for understanding and sharing human feelings. Conversely, the human characters' responses to Klara reflect their own evolving empathy towards AI, highlighting a growing recognition of AI as entities worthy of moral consideration. This mutual empathy emphasises the novel's core message, that the essence of humanity might not lie only in our genetic makeup but also in our capacity for empathy, a trait shared and reflected by beings like Klara.

In Chapter 2, the analysis centres *Machines Like Me* by Ian McEwan. This novel delves deeply into the moral conundrums and ethical quandaries that arise when humans coexist with highly advanced AI. The narrative revolved around two characters, Charlie and Miranda, and their interactions with Adam, an Artificial Companion. The chapter examines the complex relationship between these characters, focusing on the blurring of lines between human and machine, particularly in the context of empathy, morality, and physical relationships.

Adam, with human-like characteristics, serves as a prime example of the evolution of the portrayal of AI in literature. Adam's ability to perceive and navigate the world through a clear-cut ethical lens challenges the traditional view of AI as purely functional entities. Adam's character is distinctly human-like, not just in appearance but in his capacity for moral reasoning and emotional understanding. The chapter critically examines the dynamics of Adam's

relationship with Miranda. Despite his sophisticated AI design, Adam is fundamentally a machine, yet his human-like appearance plays a crucial role in the development of a physical relationship with Miranda. This relationship is fraught with complexity, as Miranda engages in sexual acts with Adam while simultaneously acknowledging and treating him as a machine. Furthermore, the chapter explores the implications of this relationship on our understanding of empathy and morality in the context of AI. It raises critical questions about the ethical boundaries in human-AI interactions and the psychological impact on humans who form deep connections with AI entities. The novel forces readers to confront the reality of AI that not only mimics human behaviour but also challenges our notions of what it means to be human.

In Chapter 3, my exploration of Marge Piercy's *He, She and It* analyses the novel's unique emphasis on AI in feminist literature, highlighting a significant thematic shift from utopian to dystopian settings in AI narratives. While previous chapters focused on Ishiguro's and McEwan's portrayal of AI as companions in more utopian contexts, such as Klara in *Klara and the Sun* and Adam in *Machines Like Me*, Piercy's narrative contrasts sharply by presenting a 21st-century dystopian world. Set in the fictional town of Tikva, meaning "hope" in Hebrew, the novel employs this setting to create an ironic juxtaposition of optimism against societal and technological challenges, with its female-centric narrative lending itself to a feminist interpretation of AI. This chapter also incorporates the theme of empathy, examining how the complex and morally ambiguous human-AI relationship between Yod and Shira enhances our understanding of empathy in literature. The novel underscores the evolving role of AI in society, and also explores how these artificial entities can evoke, mirror, and understand human emotions. This analysis of the emotional interplay between humans and AI challenges traditional views of AI as mere emotionless machines. Instead, it emphasises that the human-AI bond can be a powerful channel for emotional depth and empathetic understanding, thus reshaping our perspective on human-AI interactions. The contrast between utopian and dystopian AI narratives in literature thus offers a nuanced view of artificial intelligence, reflecting its diverse societal implications and ethical considerations.

Lastly, my final chapter focused on an analysis of two of Ted Chaing's short stories, namely "The Lifecycle of Software Objects", and "Daceys Patent Automatic Nanny". Further, I provided

a contrasting analysis of Philip K. Dicks short story, “The Mechanical Nanny”. Here, I explored the perception humans have to keep AI entities as “pets”. By exploring this relationship, this chapter shows the emotional connections between humans and digital entities, known as digients. This chapter emphasises the transformative potential of AI in forming empathetic relationships with humans, as seen by the protagonist Ana, a zookeeper who trains these digients. This aspect of the narrative reflects a significant shift in the portrayal of AI, from mere tools or companions to entities capable of forming meaningful, empathetic bonds with humans, thereby enriching our understanding of AI and its role in our lives. In “Dacey’s Patent Automatic Nanny”, I have shifted the analysis to look at the role of AI in caregiving — a theme which is similarly seen in a positive light in *Klara and the Sun*. However, in this story, the reader is shown the limitations that come alongside AI nurturing humans. It presents a view of the caregiving potential of AI, juxtaposing its advantages with the inherent constraints, which prompts a re-evaluation of our reliance on AI for such intimate and complex human tasks. This exploration broadens the scope of our understanding of AI, showing that while it can offer significant benefits in caregiving, there are essential aspects of human care and empathy that may not be fully replicable by artificial entities.

In concluding this study, it is essential to return to the foundational concepts of empathy and compassion in AI, as initially outlined with reference to Daniel Batson’s and Martha Nussbaum’s theories in my introduction. This study began with an exploration of these modalities, examining how they manifest in artificial intelligence within a literary context. Batson’s concept of empathy, defined as a motivational state generated by the welfare of others, aligns intriguingly with the portrayal of AI characters like Klara in “Klara and the Sun” and Adam in “Machines Like Me”. These AI entities, through their interactions and responses to human needs, showcase a form of empathetic engagement that resonates with Batson’s theory.

Similarly, Nussbaum’s perspective on compassion as a response to the pain of others, contingent on the recognition of shared humanity, offers a compelling lens through which to view AI in literature. The narratives in the works of Marge Piercy and Ted Chiang provide a unique exploration of compassion, challenging the reader to consider if and how AI entities might engage in compassionate behaviours under Nussbaum’s definition. This synthesis of literature

and AI reveals a profound shift in our perception of machines, not just as tools, but as entities capable of mirroring, and perhaps even understanding, human emotions within the frameworks of empathy and compassion as outlined by Batson and Nussbaum.

Looking at how the different authors in this study represent empathy and compassion in their AI characters offers an interesting perspective on human-robotic relationships. Ishiguro's *Klara and the Sun* focuses on the emotional depth and empathetic capabilities of AI, which portrays Klara as an entity capable of forming genuine emotional connections. Choosing to narrate the story from Klara's viewpoint creates a closer bond between the reader and the AI experience, thereby opening up avenues for greater empathetic understanding towards AI. This empathetic quality is similarly seen in Piercy's *He, She, and It*. Here, humans and artificial entities share a compassionate relationship - as Shira often treats Yod with respect and humility. Yod, in the same line, shows an immense sense of selflessness towards Shira as he self-destructs in order to protect her. The relationship between the digients in Chiang's "The Lifecycle of Software Objects" further depicts these entities capable of forming meaningful, empathetic bonds. In contrast, McEwan's *Machines Like Me* presents a more morally ambiguous relationship between humans and AI. McEwan highlights the ethically problems with these relationships, and forces the reader to question what makes AI entities deserving of compassion. These differing approaches provide a complex view of AI, ranging from empathetic companions to entities that challenge our very understanding of morality and ethics.

Furthermore, the portrayal of AI characters by these authors significantly influences the reader's perception and understanding of AI. Ishiguro's Klara is presented with a sense of innocence and purity in her emotional responses, which contrasts with the more morally complex portrayal of Adam in McEwan's narrative. This variation not only highlights the diverse potentialities of AI but also reflects the authors' distinct perspectives on the role and impact of technology in human life. While Ishiguro may lean towards a more harmless view of AI, integrating seamlessly into human society, McEwan presents a more cautious perspective, underlining the ethical problems and psychological impacts that advanced AI could entail.

In the time it has taken me to complete this dissertation, AI tools and technologies have made increasing inroads into our everyday lives. The use of ChatGPT and other Large Language Models has become commonplace in the personal and professional lives of many people with access to digital technologies. In 2023, a short novel entitled *I Am Code* was published, written entirely in response to human prompts by an AI called code-davinchi-002. In the education sector, tools such as Syntea have been developed to guide students in their courses. Syntea is an AI tutor that can provide feedback and verified answers to help students with their learning. The rapid advance of AI shows that an in-depth understanding of our connection with these machines is essential. Learning about these connections through works of literature can provide valuable insights into the evolving nature of this relationship, helping us to navigate the ethical, emotional, and societal implications of our increasingly AI-integrated world.

At the heart of this dissertation lies an analysis of the ways in which humans form connections with the world around them, particularly through the lens of our interactions with artificial intelligence. This journey has not only explored the evolving dynamics between humans and AI but has also delved into the fundamental nature of human relationships and emotions. As we navigate this terrain, it becomes increasingly clear how these connections, whether with fellow humans or with AI entities, shape our understanding of empathy, morality, and our humanity.

In closing, I draw on a quote by the author Kazuo Ishiguro, whose works have greatly inspired this dissertation. As Ishiguro puts it, “We all live inside bodies that will deteriorate. But when you look at human beings, they’re capable of very decent things: love, loyalty. When time is running out, they don’t care about possessions or status. They want to put things right if they’ve done wrong”. This reflection resonates deeply with the themes explored in this study. It emphasises the inherent qualities of love and loyalty that define human existence, transcending the physical limitations of our being. In the context of AI, this quote takes on a new dimension, prompting us to consider how these quintessentially human traits can be understood, mirrored, or even emulated by artificial entities.

Ishiguro’s words serve as a powerful reminder of what truly matters in the human experience — the pursuit of meaningful relationships, the act of making amends, and the capacity for profound

emotional connections. As we advance technologically, creating machines that can simulate or even understand these aspects of humanity, we are challenged to reassess our values and the essence of what it means to be human. This dissertation, therefore, aside from exploring AI in literature, serves also as a deeper inquiry into the human condition, urging us to cherish the qualities that make us human in an increasingly digital and AI-integrated world.

Reference List

Ahn, Sunyoung. “The Everyday Life of Artificial Intelligence: The Humanism of Ted Chiang’s The Lifecycle of Software Objects”. *Science Fiction Studies* 47 (2020): 73-92.

Asimov, Isaac. *I, Robot*. New York, Random House Worlds, 2004.

Atwood, Margaret. “My favourite Ishiguro: by Margaret Atwood, Ian Rankin and more.” *The Guardian*, 20 February 2021, https://www.theguardian.com/books/2021/feb/20/my-favourite-kazuo-ishiguro-margaret-atwood-ian-rankin-sarah-perry. Accessed 24 December 2023.

Avcu, Ismail. “Postmodern Ethics Revisited in Ian McEwan’s Machines Like Me: A Baumanian Reading.” *Veche*, vol. 1, no. 1, 2022, pp. 48-58.

Batson, Daniel, et al. “Is Empathic Emotion a Source of Altruistic Motivation?” *Journal of Personality and Social Psychology* 40.2 (1981): 290-302.

Batson, Daniel C. “These things called empathy: Eight related but distinct phenomena.” Decety, Jean and William Ickes. *The Social Neuroscience of Empathy*. Boston: Boston Review, 2009. 3–15.

Decety, Jean and William Ickes. *The Social Neuroscience of Empathy*. Massachusetts: Massachusetts Institute of Technology, 2009. 3-17.

Barrat, James. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York, St. Martin's Publishing Group, 2015.

Bošnjaković, Josip, and Tanja Radionov. “Empathy: Concepts, Theories and Neuroscientific Basis.” *Alcoholism and Psychiatry Research*, vol. 54, 2018, pp. 123-150.

Braidotti, Rosi. *The Posthuman*. Cambridge, Polity Press, 2013.

Cavaliere, Giulia. “Looking into the shadow: the eugenics argument in debates on reproductive technologies and practices.” *Monash Bioethics Review*, vol. 36, no. 1, 2018, pp. 1-22.

Chiang, Ted. “The Lifecycle of Software Objects.” Chiang, Ted. *Exhalation*. New York: Penguin Random House, 2019. 62-163.

---."Dacey’s Patent Automatic Nanny." Chiang, Ted. *Exhalation*. New York: Penguin Random House, 2019. 164-174.

Coeckelbergh, Mark. “Artificial Companions: Empathy and Vulnerability Mirroring in Human-Robot Relations.” *Studies in Ethics, Law, and Technology* 4.3 (2010): 1-19.

Dawson, Lesel. “Beyond Ophelia: The Anatomy of Female Melancholy.” *Lovesickness and gender in early modern English literature*, Oxford Academic, 2008, pp. 91-118.

Dick, Philip K. *Nanny*. Citadel Twilight, 1955.

---. *Do Androids Dream of Electric Sheep?* New York City: Doubleday, 1968.

DeKorver, Brittland. “Book and Media Recommendations: What If?; The Lifecycle of Software Objects; Sapiens: A Brief History of Humankind; and Gulp: Adventures on the Alimentary Canal.” *Journal of Chemical Education* (2015): 1143-1145.

Digital Dream Labs. “Purpose.” *The Butter Robot*, 30 August 2022, https://www.thebutterrobot.com/#purpose. Accessed 3 January 2024.

Erdstein, Baruch Emanuel. “Kabbalah: A Brief Definition - Kabbala is an ancient Jewish tradition integral to the Torah.” *Chabad.org*, 2023, https://www.chabad.org/kabbalah/article_cdo/aid/380617/jewish/Kabbalah-A-Brief-Definition.htm. Accessed 2 August 2023.

Fonioková, Zuzana. “The Butler's Suspicious Dignity: Unreliable Narration in Kazuo Ishiguro's The Remains of the Day.” *Brno Studies in English*, vol. 55, no. 12, 2006, pp. 87-98.

Haraway, Donna J. *A Cyborg Manifesto: Science, Technology, And Socialist - Feminism in the Late Twentieth Century*. Minnesota, University of Minnesota Press, 2016.

Hayles, Katherine N. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago, The University of Chicago Press, 1999.

Ishiguro, Kazuo. *Klara and the Sun*. London, Faber & Faber, Limited, 2021.

---. *Never Let Me Go*. London, Penguin Vintage Books, 2005.

---. *The Remains of the Day*. New York, Knopf Doubleday Publishing Group, 2010.

Keen, Suzanne. *Empathy and the Novel*. New York: Oxford University Press, 2007.

Kim, Tae Wan. "Flawed Like Us and the Starry Moral Law: A review of Ian McEwan's Machines Like Me." *Journal of Business Ethics*, vol. 170, 2021, pp. 875-879.

Malinowska, Joanna K. "What Does It Mean to Empathise with a Robot?" *Minds and Machines* 31 (2021): 361–376.

McBean, Sam. "Feminism and futurity: revisiting Marge Piercy's Woman on the Edge of Time." *Feminist Review*, vol. 107, 2014, pp. 37-56.

McEwan, Ian. *Machines Like Me*. London, Penguin Random House, 2019.

Nocks, Lisa. "The Golem: Between the Technological and the Divine." *Journal of Social and Evolutionary Systems*, vol. 21, no. 3, 1998, pp. 281-303.

Nussbaum, Martha. *Compassion: The Basic Social Emotion*. Cambridge University Press, 2009.

Patowary, Upakul. "Artificial Intelligence and Mary Shelley's Frankenstein: A Comparative Analysis of Creation, Morality and Responsibility." *Integrated Journal for Research in Arts and Humanities*, vol. 3, no. 3, 2023, pp. 121-127.

Pearson, Carol. "Women's Fantasies and Feminist Utopias." *Frontiers: A Journal of Women Studies*, vol. 2, no. 3, 1977, pp. 50-61.

Piercy, Marge. *He, She and It: A Novel*. New York, Random House Publishing Group, 2010.

---. *Woman on the Edge of Time: A Novel*. New York, Random House Publishing Group, 2010.

Rault, Jean-Loup. “Pets in the digital age: live, robot, or virtual?” *Frontiers in Veterinary Science*, vol. 2, no. 11, 2015, pp. 1-3, doi: 10.3389/fvets.2015.00011.

Roose, Kevin. “An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. (Published 2022).” *The New York Times*, 2 September 2022, https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html. Accessed 13 December 2023.

Russell, Stuart. *Human Compatible*. London, Penguin Books, 2020.

Rutherford, Adam. *Control: The Dark History and Troubling Present of Eugenics*. New York, WW Norton, 2022.

Sartre, Jean-Paul. *Being and Nothingness: An Essay on Phenomenological Ontology*. Colorado, Philosophical Library, 1956.

Sautter, Diane. “Erotic and Existential Paradoxes of the Golem: Marge Piercey's "He, She and It."” *Journal of the Fantastic in the Arts*, vol. 7, no. 2, 1995, pp. 255-268.

Scholem, Gershom. *On the Kabbalah and its symbolism*. Schocken Books, 1965, https://anarch.cc/uploads/gershom-scholem/on-the-kabbalah-and-its-symbolism.pdf. Accessed 3 August 2023.

Schülting, Sabine. “Kazuo Ishiguro’s Never Let Me Go The Role of Unreliable Narration in Redeeming Kathy’s Contested Humanity.” *Contemporary British Fiction*, pp. 1-22.

SF Dictionary. “scientifiction.” *Historical Dictionary of Science Fiction*, 1 March 2021, https://sfdictionary.com/view/104/scientifiction. Accessed 4 January 2024.

Shaviro, Steven. *Discognition*. London, Repeater, 2016.

Shelley, Mary. *Frankenstein*. London, Penguin Random House, 2006.

Shulevitz, Judith. “Review: ‘Klara and the Sun,’ by Kazuo Ishiguro.” *The Atlantic*, 2 March 2021, https://www.theatlantic.com/magazine/archive/2021/04/kazuo-ishiguro-klara-and-the-sun/618083/. Accessed 24 December 2023.

Starrs, Paul F., and Lynn Huntsinger. “The matrix, cyberpunk literature, and the apocalyptic landscapes of information technology.” *Information Technology and Libraries*, vol. 14, no. 4, 1995, pp. 251-256.

Strauss, Clara, et al. “What is compassion and how can we measure it? A review of definitions and measures.” *Clinical Psychology Review*, vol. 47, 2016, pp. 15-27.

Sterling, Bruce. Preface. *Burning Chrome*, by William Gibson. New York: Harper Collins, 1986.

Stokes, Chris. "Why the three laws of robotics do not work." *International Journal of Research in Engineering and Innovation* 2.2 (2018): 121-126.

Susen, S. (2022). Reflections on the (Post-)Human Condition: Towards New Forms of Engagement with the World? *Social Epistemology A Journal of Knowledge, Culture and Policy, 31*(6), 63-94.

Telles-Correia, Diogo, and João Gama Marques. “Melancholia before the twentieth century: fear and sorrow or partial insanity?” *Frontiers in Psychology*, vol. 6, 2015, pp. 1-4.

Turing, A. M. "I.—COMPUTING MACHINERY AND INTELLIGENCE." *MIND* LIX.236 (1950): 433-460.

Tolentino, Jia. “Ottessa Moshfegh's Painful, Funny Novel of a Young Woman's Chemical Hibernation.” *The New Yorker*, 11 July 2018, https://www.newyorker.com/books/page-turner/ottessa-moshfeghs-painful-funny-novel-of-a-young-womans-chemical-hibernation. Accessed 18 January 2024.

Unger, Arthur. “Ray Bradbury: the science of science fiction - CSMonitor.com.” *The Christian Science Monitor*, 13 November 1980, https://www.csmonitor.com/1980/1113/111356.html. Accessed 4 January 2024.

Willder, Louise. *Blurb Your Enthusiasm: A Cracking Compendium of Book Blurbs, Writing Tips, Literary Folklore and Publishing Secrets*. London, Oneworld Publications, 2022.

Wilson, Eric. *The Melancholy Android: On the Psychology of Sacred Machines*. New York, State University of New York Press, 2006.

Zimmermann, Michael E. "Authenticity, Duty, and Empathy in Do Androids Dream of Electric Sheep?" Pedersen, Hans. *Horizons of Authenticity in Phenomenology, Existentialism, and Moral Psychology: Essays in Honor of Charles Guignon*. New York City: Springer Publishing, 2014. 75-92.