



Dear, Joanne (2023) *Artificial intelligence and the Imago Dei in contemporary speculative fiction: a study of Kazuo Ishiguro's Klara and the Sun*.
MRes thesis

<http://theses.gla.ac.uk/83575/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first
obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any
format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author,
title, awarding institution and date of the thesis must be given

Enlighten: Theses
<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Artificial Intelligence and the *Imago Dei* in
Contemporary Speculative Fiction: A Study of Kazuo
Ishiguro's *Klara and the Sun*

Joanne Dear MA (Hons), MSc

Submitted in fulfilment of the requirements for the Degree of
Theology and Religious Studies, MRes

Theology and Religious Studies
School of Critical Studies
University of Glasgow

September 2022

Abstract

This thesis critically examines understanding of the theological concept of the *imago Dei* in relation to artificial intelligence (AI) using Kazuo Ishiguro's speculative novel *Klara and the Sun* to do so. The study finds that this genre of fiction is a useful tool for examining hypothetical scenarios, in this instance the arrival of Artificial General Intelligence or 'strong-AI', owing to its ability to consider unverifiable realities. Using textual analysis, this research determines that examining human uniqueness from a theological perspective in consideration of AI and robotics is not a straightforward discourse. The substantive model within the novel is largely discounted owing to the superior intellectual capacities demonstrated by the AI presented, with the caveat that human and machine thinking are two very different processes. Klara, the narrator of the novel and an AI, fits into the functional interpretation well as she is a product of human activity in the created world, with the concerns this model historically generates around the exploitation of nature and the replacement of human beings demonstrated. The relational view of the *imago Dei* finds Klara with the ability to engage in AI-AI and AI-human relationships, and with the creation of her own religion she exhibits the nascent possibilities of an AI-divine relationship. Klara's external motivation to protect the child she is purchased for can be read in terms of Pannenberg's idea of exocentricity, found in the eschatological model of the image. The paper concludes that the AI presented by Ishiguro may be made in the *imago humanitatis* with a 'weak-imaging' connection to the *imago Dei*.

Table of Contents

Acknowledgements	1
Author's Declaration	2
Abbreviations.....	3
Introduction	4
Chapter One: Literature and Theology as Methodology.....	6
Speculative Fiction.....	11
Chapter Two: The <i>Imago Dei</i>	17
Biblical <i>Imago Dei</i>	18
The Substantive Interpretation.....	20
The Functional Interpretation.....	22
The Relational Interpretation	24
Created Co-Creators?	29
Eschatological <i>Imago Dei</i> ; A Fourth Way?	30
Chapter Three: AI and Literature	33
A Brief History of AI.....	34
A Brief History of AI in Literature.....	36
Kazuo Ishiguro and <i>Klara and the Sun</i>	39
Summary of <i>Klara and the Sun</i>	40
<i>Klara and the Sun</i> as Speculative Fiction	42
Klara as Narrator	44
Chapter Four: AI and the <i>Imago Dei</i> in <i>Klara and the Sun</i>	47
Klara as AI	47
The Question of Consciousness	49
The Substantive Model and Klara	51
The Functional Model and Klara.....	54
The Relational Model and Klara	57
The Eschatological Model and Klara	61
A New Member of the Family?.....	63
Conclusion	65
Key Findings.....	65

Limitations	67
Areas for Further Research	68
Bibliography	69

Acknowledgements

With many thanks to my supervisor, Professor Heather Walton, whose guidance and suggestions have been invaluable. Having begun this part-time MRes in 2017, I would like to express my gratitude to Julie Clague for her support in encouraging me to return to complete it after two wonderful years at home with my daughter. For their love and encouragement throughout this process, as always, my deepest thanks are extended to Stephen and Poppy.

Author's Declaration

I certify that this thesis is my own work, except where indicated by referencing, in accordance with University guidelines.

Abbreviations

AF	Artificial Friend
AGI	Artificial General Intelligence
AI	Artificial Intelligence
MIT	Massachusetts Institute of Technology
SF	Science Fiction

Introduction

The doctrine of the *imago Dei*, that human beings bear the image and likeness of God, is one of the central tenets of Christian theological anthropology. It supports understanding of the place, function, and destiny of human beings within the created order. In this dissertation I will explore the challenges, opportunities, and insights the emergence of truly intelligent artificial intelligence (AI) could bring to the image. As this is currently a hypothetical situation, a future possibility, literature will be used as the vehicle for exploration. Kazuo Ishiguro's *Klara and the Sun* is the novel that has been chosen for examination, a piece of speculative fiction narrated by a highly intelligent, sophisticated, and self-aware robot. The novel will act as a case study, or thought experiment, through which the intersection of the *imago Dei* and AI can be considered, asking questions such as if AI can become indistinguishable from a person, what does this mean for human distinctiveness? Would theological anthropology still be able to consider human beings as made in the image? Is there possibly a way that the study of AI in this context can help clarify rather than challenge the idea of human uniqueness?

Chapter One explores what is meant by the term 'literature and theology' and will examine the interdisciplinary writings of N.Katherine Hayles, Katie Cannon, Martha Nussbaum and Cynthia Wallace to support the formulation of a methodology for this dissertation. Key to this process will be ensuring a balance is maintained between the two fields of study, with literature and theology intertwining rather than one dominating the other. This chapter will also consider the category of speculative fiction, including a discussion of how the term can be defined and why the genre is suitable to lend itself to theological reflection.

Chapter Two will examine historical and contemporary understandings of the *imago Dei*. This will begin with a brief overview of the motif in biblical scripture before examining the strengths and weaknesses of the substantive, functional, relational and eschatological models, with an emphasis on interpretations that are particularly relevant within the context of AI. The work of the computer scientist and theologian Noreen Herzfeld, who has published widely on the intersection of

theology, science, and technology, will form the focus of this section. In her writing, Herzfeld examines historical shifts in understanding of the image and considers how these are reflected in comparable changes in approaches to the field of AI and robotics research. She also raises important points concerning the ethical and relational considerations AI brings to the *imago Dei*, including reflecting on why human beings pursue the creation of advanced robots and what it means for us to be made in the image of God in a world altered by AI and technology.

Chapter Three will outline a brief history of AI, from its origins in the automata of the past to the possible realisation of ‘Artificial General Intelligence’ (AGI) in the future, a prospect in which AI would be able to understand or learn any intellectual task currently only attainable by a human being. A short survey of AI in literature will follow, examining examples of those who were ‘made’ rather than born in stories from Greek myth to contemporary speculative fiction. Discussion will then turn to *Klara and the Sun*, where I will establish my justification for choosing this particular text, analyse its status as a work of speculative fiction and summarise the novel’s plot.

Chapter Four will textually analyse *Klara and the Sun* in detail and examine what Ishiguro’s created world tells us about how the AI presented may connect to the idea of the *imago Dei*. This will include considering if the AI he poses qualifies as AGI and if so, are they also in possession of consciousness? Other questions that will be contemplated include could AI preserve a human life in the event of their biological death? Is it possible for a person’s agency, consciousness, and memories to be uploaded and continued in a robotic? In whose image is this AI made? All of these points will be considered in context of the four models of the *imago Dei*, substantive, functional, relational and eschatological, discussed in Chapter Two.

The theological discourse concerning AI is still a relatively emerging discipline with ample opportunity to add to its body of work. It is hoped that this dissertation will contribute to this growing field as well as the more established area of literature and theology.

Chapter One: Literature and Theology as Methodology

This chapter opens with an exploration of what we mean by the term ‘literature and theology’, before turning to examine the methods others have adopted in relevant interdisciplinary work. A review of the approaches engaged by N.Katherine Hayles, Katie Cannon, Martha Nussbaum and Cynthia Wallace will be considered, and their viewpoints on what, if anything, literature can tell us about reality will also be investigated. This project intends to use speculative fiction as a vehicle for theological exploration so a discussion of the genre then follows, considering definitions, why it may lend itself to such reflection, and concludes with the identification of any limitations this approach may have.

Theologian and literary scholar David Jasper writes that ‘theology simply cannot be studied alone inasmuch as it participates in the complex unity of which it is a part and which is alone accessible through the broad exchanges made within culture.’¹ A good starting point for defining the term ‘literature and theology’ is Heather Walton’s description of the field as a ‘liminal meeting space where new possibilities can take shape.’² As Walton continues, this is certainly not an area without its challenges and disagreements, with blurred boundaries and tensions sometimes present between the disciplines. As late as the 1970s, academic study in the interdisciplinary field of literature and theology generally ‘looked back with a degree of religious nostalgia and a sense that things still could be as they always seemingly had been.’³ In many ways this world, where literature highlighted the ‘truths’ of religion and tended to focus on Christian practice and belief, was represented by T.S. Eliot’s requirement for the necessity of ‘Christian readers to scrutinize their reading, especially of works of imagination, with explicit ethical and theological standards.’⁴

¹ Jasper, D. *Literature and Theology as a Grammar of Assent*, (Farnham: Ashgate, 2016), 7.

² Walton, H. ‘Introduction’ in Heather Walton (ed.), *Literature and Theology: New Interdisciplinary Spaces*, (Farnham: Ashgate, 2011), 2.

³ Jasper, D. ‘Interdisciplinarity in Impossible Times: Studying Religion through Literature and the Arts’ in Heather Walton (ed.), *Literature and Theology: New Interdisciplinary Spaces*, (Farnham: Ashgate, 2011), 5.

⁴ Eliot, T.S. ‘Religion and Literature’ in *Selected Essays*, 3rd edition (London: Faber and Faber, 1951), 388.

In his reflections on literature and theology, Andrew Hass first examines the belief that we imagine that we ‘know’, that is, we imagine we know ‘the structures of knowledge, the systems and disciplines we need to accrue knowledge’⁵ even if we are aware some knowledge itself will evade us. When addressing the disciplines of literature and theology from a historical perspective, Hass finds worlds that have ‘been safe - safe because necessary - to imagine we know.’⁶ Literature, he proposes, lands us in a world created by the imagination of both the text and the reader, or in an ‘imaginary state of knowing.’⁷ Theology on the other hand imagines our knowledge of the unseen can be powerfully rooted, even without empirical evidence. To Hass, it is not that God is simply an invention of our imagination, but that our knowledge of God becomes ‘imagined and imaged as real - God as the Real. What we know is an image of the Real.’ Hass links this idea to the tradition of the *imago Dei*, where our own creative worlds reflect God’s. He concludes that we could imagine literature and theology not as distinct categories shaped into an ‘esemplastic unity’⁸ but instead as a way of surpassing distinct systems of knowledge, essentially widening out the creative dialogue beyond discrete categories and allowing it to become a vehicle for a new way ‘of thinking, of thinking in disciplines, of thinking imagination, of thinking how we imagine that we know and what we know?’⁹

Literature has been used, in this frame, as a lens through which to study concepts, themes, and issues within the discipline of theology, with its focus broadening from Christian practice and belief into fields including feminism, gender and race. Relevant to this dissertation is the interdisciplinary work of N. Katherine Hayles, who raises questions about human nature that originate in theological traditions, including examination of the moral status of non-human beings. Her concept of the posthuman is framed by the proposition that we have already reached the point of ‘posthumanity’, the important question is what kind of posthumans will we be? Hayles acknowledges the complexity of her idea, writing that ‘it involves a range

⁵ Hass, A. ‘Discipline Beyond Disciplines’ in Heather Walton (ed.), *Literature and Theology: New Interdisciplinary Spaces*, (Farnham: Ashgate, 2011), 19.

⁶ Ibid

⁷ Ibid

⁸ Ibid, 20.

⁹ Ibid

of cultural and technical sites'¹⁰ including AI and artificial life. She details some aspects of posthumanism and reports that:

the posthuman view privileges informational pattern over material instantiation...the posthuman considers consciousness ... as an evolutionary upstart trying to claim that it is the whole show when in actuality it is only a minor sideshow...the posthuman view thinks of the body as the original prosthesis we all learn to manipulate...and the posthuman view configures human being so that it can be seamlessly articulated with intelligent machines.¹¹

The properties of the posthuman Hayles describes here will make for useful analysis in the text that will come under consideration in this thesis, whose eponymous narrator is an advanced AI robotic. Part of my inspiration for choosing to examine *Klara and the Sun* was Hayles's own analysis of Greg Egan's fiction.¹² She discusses his 'hard' SF novels in relation to many of the ideas raised in *How We Became Posthuman*, including questions around the uploading of human consciousness to computers and the prospect of a post-biological future for humankind. These subjects are also considered in Ishiguro's novel and whilst a very different author to Egan, they both envisage the human-computer future relationship not as 'a question of technology...but as an ontological inquiry into the relation of humans to the universe.'¹³ Hayles differs from Donna Haraway in the anxiety she demonstrates concerning how posthumanism may present itself; there is apprehension as well as optimism about human technological power. Haraway states 'at the centre of my ironic faith, my blasphemy, is the image of the cyborg'¹⁴. She finds that the cyborg challenges key approaches to the human that are rooted in theology, questioning the anthropocentric focus of Christianity as well as the emphasis on nature as being in opposition to technology.

Following a methodology that links literature and theology also makes provision to consider the question: what can literature tell us about reality? There is range of

¹⁰ Hayles, N.K. *How we Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, (University of Chicago, 1999), 247.

¹¹ *Ibid*, 2-3.

¹² Hayles, N.K. *My Mother was a Computer: Digital Subjects and Literary Texts*, (University of Chicago, 2005), 215-240.

¹³ *Ibid*, 218.

¹⁴ Haraway, D. 'Cyborg Manifesto', in *Simians, Cyborgs and Women: The Reinvention of Nature*, (Taylor and Francis Group, 1991), 149.

thought about what, if anything, literature can convey about real life. Katie Cannon's interdisciplinary work between literature and theology, often focusing on the writing of black women, has an unproblematised reliance upon reading literature as reality. Cannon moves away from the idea of literature as an expression of creativity or art, to a view where literature is a way of obtaining knowledge of real life. A criticism of Cannon's call for literature to be read as reality is that this encourages readers to only engage with literature that confirms already entrenched beliefs and ideas, whereas the influential philosopher Martha Nussbaum moves away from this notion, believing literature should challenge our beliefs and perhaps even change us.

Nussbaum has reflected on the relationship between literature and philosophy, focusing on the former as a moral guide rather than a revelatory tool. She examines the contribution made by texts to the exploration of important questions about human beings and human life, writing those truths can only be 'fittingly and accurately stated in the language and forms characteristic of the narrative artist.'¹⁵ Nussbaum argues that moral philosophy and moral thinking have an absolute need for literature, its significance comes from its ability to raise important points about human lives and situations that may otherwise be overlooked, permitting us to build through imagination 'a bridge that allows the other to become an object of our compassion.'¹⁶ Additionally, Nussbaum notes that in some readers, these works can evoke high levels of reflective consciousness that can move us closer towards truth than any other method. She finds value to moral philosophy presented in literature as it paints a believable picture of life, while examining the important motifs of love, friendship, death and so on. Walton highlights that Nussbaum's:

Advocacy of literature is the acknowledgement that there may be some views of the world, particularly those which emphasise the beauty, mystery, complexity and ambiguity of existence, that cannot be fully and adequately stated in the language of conventional philosophical prose.¹⁷

¹⁵ Nussbaum, M. *Love's Knowledge: Essays on Philosophy and Literature*, (Oxford: Oxford University Press, 1990), 5.

¹⁶ Nussbaum, M. *Upheavals of Thought: The Intelligence of Emotions*, (Cambridge University Press, 2001), 66.

¹⁷ Walton, H. *Literature, Theology and Feminism*, (Manchester: Manchester University Press, 2007), 17.

However, Walton also makes the important point that this view turns literature into a ‘a necessary supplement to philosophy rather than an equal partner.’¹⁸

Cynthia Wallace highlights the importance of not viewing literature as being ‘used’ by theology in any way, but finding its ‘potential to do good in the realm of material justice strongest when it is attended to as *literature*, when it is read closely, not “used” for some purpose but given open attention, with a heightened awareness to its singularity.’¹⁹ Wallace’s idea that literature is not a means by which we can confirm or find ‘proof’ for already entrenched theological standpoints is very important, instead it should say something about reality and also ‘mean something about what it is to mean anything at all...[the] duality of literature as both a space for presenting things as they are and for hinting at the imaginative potential of how things could be.’²⁰

Having examined the methodologies detailed above, I would conclude that one of the biggest challenges to working in an interdisciplinary manner is retaining balance between the two disciplines. I aim for literature and theology to be equal partners in this study and will evaluate the success, or otherwise, of my approach at the end of this dissertation. I hope that following an interdisciplinary approach will allow an informed exploration of how literature may confirm or challenge the established theories of the *imago Dei* discussed in Chapter Two. While there is much to admire in Cannon’s work, I disagree with her view that literature should be read as reality, believing that Wallace’s dance between showing things as they are and demonstrating the ‘imaginative potential’²¹ of storytelling to be more akin to my own point of view. I am intrigued by the notion that complex theological or philosophical ideas can be made much more accessible and engaging through the medium of literature, *Klara and the Sun* for example navigates deep ontological questions but the skilful and deft storytelling allow this to be done in a novel that at first glance may seem quite simple, almost fable like.

¹⁸ Ibid, 18.

¹⁹ Wallace, C. *Of Women Borne: A Literary Ethics of Suffering*, (New York: Columbia University Press, 2016), 34.

²⁰ Ibid, 215.

²¹ Ibid.

Speculative novels such as *Klara and the Sun* project the reader forward into a potential future, allowing us to explore possibilities and ask deep questions about situations that can only be imagined just now but may well be the moral and ethical challenges of tomorrow. Attention will now turn to examining this genre in more detail and assessing why it may be a useful vehicle for examining theological questions and concepts.

Speculative Fiction

While all fiction, in a very literal sense, can be considered speculative in nature, there is a long history of this type of storytelling that runs from ancient times to present day, including the myth of Atlantis found in Plato's dialogues *Timaeus* and *Critias*, Thomas More's *Utopia*, Aldous Huxley's *Brave New World* and George Orwell's *1984*. There are a handful of references to the term 'speculative fiction' towards the end of the nineteenth-century, but the first use of the term is generally attributed to Robert A. Heinlein in 1947, who stated that 'speculative fiction (I prefer that term to science fiction) is also concerned with sociology, psychology, esoteric aspects of biology, impact of terrestrial culture on the other cultures we may encounter when we conquer space etc.'²² The high level of prescription in Heinlein's definition is perhaps useful for identifying or defining 'quality' science fiction (SF), but it began to fall out of use in the 1960s, eventually drawing criticism from figures such as the literary critic and author Samuel Delany. Usage of the term itself evolved and was embraced by several traditions in the SF field, including Judith Merril's work within feminist speculative fiction in the 1970s. She noted that the genre is 'a special sort of contemporary writing which makes use of fantastic and inventive elements to comment on, or speculate about, society, humanity, life, the cosmos, reality and any other topic under the general heading of philosophy.'²³

In 1975, Robert Scholes published *Structural Fabulation: An Essay on the Fiction of the Future* in which he emphasised that contemporary literature should be principally occupied with fictional explorations of human situations influenced by science, or its consequences, in order to support readers to 'break the circle of

²² Heinlein, R.A. and Heinlein, V. *Grumbles from the Grave*, (Del Rey, 1989), 49.

²³ Merril, J. *Science Fiction: The Best of the Best*, (New York: Delacorte Press, 1967), 3.

indifference and act in accordance with a structural perception of the universe.’²⁴ Scholes highlights that within the primary forms of post-Enlightenment fiction, realism is the dominant tradition whereas SF, fantasy and other non-mimetic genres are threads of the subversive and varied ‘fictional form that is both old and new, rooted in the past but distinctly modern, oriented to the future but not bounded by it,’²⁵ a form Scholes calls fabulation. He pinpoints three ideologically and historically defined forms, or ‘mutations’, of fabulation: the dogmatic, which includes works such as John Milton’s *Paradise Lost* and is usually underpinned by a religious standpoint, the speculative, which includes works such as Jonathan Swift’s *Gulliver’s Travels* and is normally rooted in humanist ideas. And lastly the structural, which appeared at the beginning of the twentieth century and utilised scientific concepts such as Darwin’s theory of evolution and Einstein’s theory of relativity. Despite the importance of the rise of science and focus on what we *know* about the world in the development of structural fabulation, Scholes emphasises that both human and physical sciences are drawn upon equally. Diana Waggoner, an academic whose focus is largely on fantasy literature, argues to the contrary, proffering that it is only what we do not know that we can speculate on. She theorises that speculative fiction began to emerge at the end of the nineteenth century as a broad, multi-genre category that addressed the ‘supernatural and/or non-existent phenomena (such as the future) as a special class of objectively real things or events.’²⁶ For Waggoner, the development of speculative fiction allowed for a genre in which writing that was otherwise realist in nature could also consider unverifiable realities.

Margaret Atwood has written that SF and speculative fiction are fluid terms, with her thoughts on the genre generally supported by professor of education Paul L. Thomas in his work in this area.²⁷ She finds the distinction between SF and speculative fiction lies in possibility, stating that the latter can be defined by ‘plots that descend from Jules Verne’s books about submarines and balloon travel

²⁴ Scholes, R. *Structured Fabulation: An Essay on the Fiction of the Future*, (Norte Dame and London: University of Notre Dame Press, 1975), 1.

²⁵ *Ibid*, 27.

²⁶ Waggoner, D. *The Hills of Faraway: A Guide to Fantasy*, (New York: Atheneum, 1978), 9.

²⁷ See: Thomas, P.L. (Ed.) *Science Fiction and Speculative Fiction: Challenging Genres* (Sense Publishers, 2013).

and such - things that really could happen but hadn't completely happened when the authors wrote the book.'²⁸ Atwood built on Judith Merril's work and by the late 1980s she began to describe some of her dystopian novels as speculative in nature, arguing against classifying *The Handmaid's Tale* as SF because it 'contains no intergalactic space travel, no teleportation, no Martians.'²⁹ One difficulty with this definition is where, and how, do we draw the boundaries between what is possible and impossible? Whereas Atwood classifies stories about interstellar travel as SF because she believes it will never transpire, others may well disagree. This issue is the problematic crux of Atwood's definition and somewhat boxes us in to a limited and restricted perception of our reality, or our potential reality.

Writing in 2017, Marek Oziewicz identified three historically located meanings of the term 'speculative fiction'; as a sub-genre of SF that deals with human rather than technological problems, as a genre distinct from and opposite to SF in its exclusive focus on possible future, and a super category for all genres that deliberately depart from imitating the 'consensus reality'³⁰ of everyday experience. Oziewicz observed an emerging viewpoint that speculative fiction refers to 'a fuzzy set super category that houses all non-mimetic genres'³¹ and goes on to embrace Pierre Bourdieu's notion of a 'cultural field' to describe the classification. Oziewicz writes that the functions of speculative fiction include its use as a device to 'dismantle the traditional Western cultural bias in favour of literature imitating reality...and it is a quest for the recovery of the sense of wonder across its semantic spectrum.'³² While the term speculative fiction has only recently begun to make headway in academic circles, Oziewicz points out that younger readers and authors, as well as online forums and resources, wholeheartedly own the label of speculative fiction as a way to 'conceptualise its experience of new types of non-mimetic writing and to position them in a continuous relation to older, ideologically loaded forms.'³³ Oziewicz highlights the

²⁸ Atwood, M. *In Other Worlds: SF and the Human Imagination*, (Virago, 2012), 6.

²⁹ Atwood, M. *Writing with intent: Essays, reviews, personal prose: 1983-2005*, (New York: Carroll and Graf Publishers, 2005), 285.

³⁰ Oziewicz, M. 'Speculative Fiction' in *Oxford Research Encyclopaedia of Literature*, (Oxford University Press, 2017), 1.

³¹ *Ibid*

³² *Ibid*, 3.

³³ *Ibid*, 7.

inclusiveness and open-endedness of speculative fiction as a much appreciated advantage of the genre, defining itself by what it includes rather than what it excludes and ‘for those who value the term, it is the largest, the most diverse, and the most dynamic category of modern storytelling.’³⁴

Examining the work of Merril, Scholes, Waggoner, Atwood and Oziewicz demonstrates the broadness of the term ‘speculative fiction’. Harry Eiss highlights this variety in his reference to it as ‘the umbrella phrase for fiction that projects possibilities beyond the present literal world.’³⁵ Speculative fiction is therefore a nebulous term for an over-arching genre, it may include SF and fantasy, but also all the genres that intersect with these primary categories, such as alternate history, horror, utopian and post-dystopian literature. While Heinlein’s definition has fallen out of fashion, his core idea that speculative fiction acts as a synonym for science fiction remains relatively unopposed, especially outside of academic circles.

Having examined how speculative fiction as a category can be defined, consideration will now turn to how a novel from the genre, Ishiguro’s *Klara and the Sun*, may be useful in investigating AI’s interaction with theological representations of the *imago Dei*. Merril writes that speculative stories allow the exploration and discovery by means of projection, extrapolation, analogue, hypothesis-and-paper-experimentation, something about the nature of the universe, of man, of ‘reality’.³⁶ This suggests the genre is a suitable vehicle for the exploration of ontological questions, including reflection on the distinction between machine and human as technology advances, an idea at the crux of *Klara and the Sun*. The ‘nascent anxieties’³⁷ Elaine Graham writes about concerning human apprehension over the future are deeply intertwined with ontological questions about the principles of human identity and the boundaries between the human and non-human, and the complications of being created beings who are also

³⁴ Ibid, 19.

³⁵ Eiss, H. (ed). *Electric Sheep Slouching Towards Bethlehem: Speculative Fiction in a Post Modern World*, (Cambridge Scholars Publisher, 2004), cvii.

³⁶ Merril, J. ‘What Do You Mean: Science? Fiction?’ in Thomas D. Clareson (ed). *Science Fiction: The Other Side of Realism*, (Bowling Green, OH: Bowling Green University Popular Press, 1971), 60.

³⁷ Elaine Graham, *Representations of the Post/human*, 14.

creators. This is an idea I will return to again in Chapter Two in discussion of Philip Hefner's notion of 'created co-creators.'

Speculative novels can also function as thought experiments, imaginary worlds that may help us learn something new about reality. In their work on the epistemic value of the genre, Johan De Smedt and Helen De Cruz write that analytic philosophy, especially thought experiments, and speculative fiction 'rely on similar cognitive mechanisms'³⁸ including considering future possibilities and imagining potential possibilities. Speculative novels tend to ask the question 'what if?' about a potential situation, then propose an answer. For example, Philip K. Dick's *The Man in the High Castle* asks: 'what if the Allies had lost World War Two leading to German and Japanese occupation of most of the USA?' Ursula K. Le Guin's *The Left Hand of Darkness* asks: 'what if we lived in a genderless society where given genders could be temporarily assumed depending on whom we are in a relationship with?' One of the questions Ishiguro asks in *Klara and the Sun* is: 'what if we lived in a society where AI and robotics research had advanced to a point where it could carry out almost any human task?'

It is hoped that viewing *Klara and the Sun* as a thought experiment will be a valuable way of examining the novel, but there are limitations to what this can tell us that should be considered from the outset. What makes speculative novels exciting to me as a reader is that they probe our reality, looking at what might be or what might have been, allowing us to read and re-read the world in innumerable ways, but it is important to state that no claims about what is 'real' or factual can be made despite the fact the genre has the potential for 'challenging consensus reality...making speculative fiction politically scrappy, cognitively empowering, and affectively stimulating.'³⁹

The purpose of this chapter was to review the methodologies of those who have sought dialogue between literature and theology before and identify an approach to my own work. Attention will now turn to examine the doctrine of the *imago Dei*,

³⁸ De Smedt, Johan. and De Cruz, Helen. (2015). The Epistemic Value of Speculative Fiction in *Midwest Studies in Philosophy*, (39, 1), 59.

³⁹ Oziewicz, *Speculative Fiction*, 19.

a central theme in Christian anthropology, to understand more fully the place of artificial intelligence and humanity within the world.

Chapter Two: The *Imago Dei*

The doctrine of *imago Dei*, that human beings bear the image and likeness of God, is one of the central tenets of Christian theological anthropology. Despite this prominence, there is little agreement between theologians on what the term actually means. J. Wentzel van Huyssteen describes the image as ‘a set of complex theological mini-theories’⁴⁰ rather than one single unified approach, and Jeanine Thewatt-Bates writes that there is a ‘plasticity’⁴¹ to the term, citing biblical ambiguity and the ways in which contemporary issues have shaped understanding as the reason for this. Interpretations of the image have varied throughout history; the substantive model looks for a quality or set of qualities in the human, such as reason, rationality or intellect, whereas the functional view finds the image embedded in human actions or our maintenance and enforcement of dominion over Earth. Theologians including Karl Barth see the image as developing in the inter-relationship of two beings, either human with human or human with divine.

As a detailed survey of what each theologian has to say on the *imago Dei* is outside the remit of this dissertation, this chapter will concentrate on historical and contemporary interpretations in particular reference to the context of AI, and how they limit or expand upon and apply understanding of the image of God in this setting. This study will concentrate on Christian interpretations of the image, although the concept is present elsewhere within world religions. Key figures who have published on the intersection of theology and robotics include the computer scientists and theologians Anne Foerst and Noreen Herzfeld. Foerst views AI and robotics as a means for human self-understanding and a thought-provoking position from which to consider the *imago Dei*. Her research focuses on the work of Paul Tillich, oral and scripture traditions, and the importance of human emotion.⁴² Herzfeld builds on Foerst’s ideas and also includes the challenging ethical and relational questions AI brings to the image. Herzfeld employs an intersectional

⁴⁰ van Huyssteen, J.W. *Alone in the World? Human Uniqueness in Science and Theology*, The Gifford Lectures (Grand Rapids, MI, 2006), 313-14.

⁴¹ Thewatt-Bates, J. *Cyborg Selves: A Theological Anthropology of the Posthuman*, (Routledge, 2012), 109.

⁴² See for example: Foerst, A. *God in the Machine: What Robots Teach Us About Humanity and God*, (New York: Dutton, 2004) and Foerst, A. (1996). *Artificial Intelligence: Walking the Boundary*, *Zygon*, 31:4, pp.681-693.

approach between theology, science and technology to examine historical shifts in understanding and interpretation of the *imago Dei* and how this is reflected in comparable changes in approaches to AI and robotics research. She raises important questions in her research, including why do humans continue to pursue the possibility of creating a machine in our own image⁴³ and what does it mean for humans to be made in the *imago Dei* in a world altered by robotics and AI? In order to consider the full range of Herzfeld's writing on this topic her work will form the focus of this dissertation.

Discussion of the *imago Dei* will begin with a brief overview of the concept in scripture, before moving on to examine the substantive, functional and relational models in relation to what extent they mirror developments in AI research. Attention will then be turned to the idea of 'created co-creators', a concept that has roots in both the functional and relational view. Finally, the eschatological interpretation of the *imago Dei* will be considered as a fourth option, a model which may have particular relevance to the field of AI and robotics.

Biblical *Imago Dei*

The *imago Dei* is a foundational concept in Christian understanding of human nature. In Genesis 1:26-28⁴⁴ it is revealed that God intentionally created humankind in Their own image and after Their likeness. Following this, a mandate is granted for humanity to act within and upon the created order. The biblical passages in question are quoted in English translation and in full:

²⁶ And God said, Let us make man in our image, after our likeness: and let them have dominion over the fish of the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creepeth upon the earth.

²⁷ So God created man in his own image, in the image of God created he him; male and female created he them.

²⁸ And God blessed them, and God said unto them, Be fruitful, and multiply, and replenish the earth, and subdue it: and have dominion over the fish of the

⁴³ Herzfeld, N. *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis: Fortress Press, 2002), 5.

⁴⁴ All biblical references refer to the King James Version (KJV) of the Bible.

sea, and over the fowl of the air, and over every living thing that moveth upon the earth.

These verses are at the core of *imago Dei*, the idea that human beings are in possession of the image and likeness of God and that this position somehow confers a distinctive place for humanity above the rest of creation. The book of Genesis makes two further references to the image, firstly in Gen 5:1-3 with a description of the relationships between God, Adam, and Seth in which God's creation of humanity is reiterated and secondly in Gen 9:6, where the creation of humanity in the *imago Dei* is related to the sanctity of human life: 'Whoso sheddeth man's blood, by man shall his blood be shed: for in the image of God made he man.'

Many of the references to the *imago Dei* in the New Testament are found in writings attributed to Paul, with passages including 2 Corinthians 4:4 and Colossians 1:15 focusing on Christ as the perfect and true image of God. The image in the New Testament demonstrates the harmony that exists between Christ and God, and the discord between creator and creature. Resolution can be found through Christ, with 2 Corinthians 3:18 revealing:

¹⁸ But we all, with open face beholding as in a glass the glory of the Lord, are changed into the same image from glory to glory, even as by the Spirit of the Lord.

From an eschatological perspective, 1 Corinthians 15:49 informs us that ultimately 'as we have borne the image of the earthly, we shall also bear the image of the heavenly'. The New Testament's repositioning of considerations on creation and redemption with the image motif in a Christological sense will be explored later in regards to the work of Karl Barth.

The *imago Dei* is a significant theological concept but given its importance, its rarity and brevity in scripture is noticeable. Herzfeld find scripture offers little in the way of succinct definition of the image, writing that 'the problem with Genesis 1 is that it does not describe what this image is. Nor is the image described anywhere else in the Christian scriptures.'⁴⁵ The lack of a firm definition has historically raised several questions including what, if anything, is the difference

⁴⁵ Herzfeld, N. (2007). A New Member of the Family? The Continuum of Being, Artificial Intelligence, and the Image of God, *Theology and Science*, 5:3, 238.

between the terms ‘image’ and ‘likeness’? And how should we define them? Why does God speak in the plurals ‘us’ and ‘our’? Attempts to answer these questions have led to the emergence of three primary ways of viewing the *imago Dei* in Christian theology; the substantive, where the image is held in a particular quality within the individual, the functional, where the image is seen in action, and the relational, where the image is found in the relationships we have with God and each other.

The Substantive Interpretation

The substantive model is a classical and influential interpretation of the *imago Dei*, referring to the substance of God which humans share. The image in this understanding is most often associated with reason, rationality and intellect, human capacities that animals are not usually believed to possess⁴⁶. Saint Augustine’s writing dominated thought on the *imago Dei* in the medieval world and beyond, with the belief that humans bear God’s image from their creation but the image remains unfulfilled, the cause of sin. He supported the idea of the image as present in a person’s metaphysical substance, the mind or the soul, asserting that evidence of God could be found in the pinnacle of humankind, the mind, proposing that ‘we must find in the soul of man, i.e., the rational or intellectual soul, that image of the Creator which is immortally implanted in its immortality.’⁴⁷ Scottish theologian David Cairns writes that ‘in all the Christian writers up to Aquinas we find the image of God conceived of as man’s power of reason’.⁴⁸ While this may be somewhat of an overstatement on Cairns’s part, it is the case in the substantive model that *some* individual human quality or qualities exist within our nature in which the image resides. Herzfeld notes that in this interpretation, ‘the divine image, as a human quality, becomes part of the substance of our very being.’⁴⁹

Substantive approaches to the *imago Dei* initially dominated theology’s response to AI, but more recent explorations have departed from this view. Criticism of the

⁴⁶ In Cortez, M. *Theological Anthropology: A Guide for the Perplexed*, (T.&T. Clark Ltd, 2009), 19. the author notes that humans are also animals and any characteristics or capacities we can identify as uniquely human can be found in other animals.

⁴⁷ Augustine, *Confessions*, (London: Penguin, 1961), xiii.

⁴⁸ Cairns, D. *The Image of God in Man*, (New York: Philosophical Library, 1953), 60.

⁴⁹ Herzfeld, *In Our Image*, 16.

model includes its ranking of creation into a hierarchy, raising concerns around the treatment of women, animals and the environment itself. In terms of sex and gender, van Huyssteen highlights these issues, writing that ‘the image of God was very closely and most directly connected with males, who were allegedly better ruled by reason, while females were tied to their embodiedness.’⁵⁰ When considering other living animals, David Fergusson finds that the *imago Dei* ‘does not enable...some shortcut to identifying a single property or function that differentiates us from the other animals.’⁵¹ I feel that the substantive model would ‘other’ AI in a similar way to the rest of non-human creation and I have difficulties with giving credence to a model that separates humankind from all other forms of creation to emphasise our uniqueness. Indeed, the development of truly strong AI would be a major setback to the substantive model; if technology could outperform human beings at all tasks requiring rationality, where would that leave human uniqueness and distinctiveness? For Herzfeld, the substantive model is also inadequate because it does not consider the importance of relationship in shaping human intelligence⁵² and views the interpretation as an undesirable attempt to ‘escape from the messiness of human physicality’.⁵³

The substantive view has largely been rejected by modern Christian thought, although in contemporary theology Tillich interpreted the *imago Dei* as rationality combined with a factor he called ontological reason⁵⁴, which allows us to understand ‘complex levels of reality, giving human beings a wholeness and a way of perceiving wholeness that other animals lack.’⁵⁵ However, Tillich’s understanding of the image also has roots in the relational interpretation, linking relationships with each other or with God to his philosophy of action. Within the substantive outlook Marc Cortez identifies the ‘glaring weakness’⁵⁶ that is the lack of exegetical support for the model. Herzfeld writes that current theological

⁵⁰ van Huyssteen, *Alone in the World?* 127.

⁵¹ Fergusson, D. (2013). Humans Created According to the Imago Dei: An Alternative Proposal, *Zygon*, 48:2, 449.

⁵² Herzfeld, N. (2004). Creating in Our Own Image: Artificial Intelligence and the Image of God, *Zygon*, 37:2, 305.

⁵³ Herzfeld, N. *Technology and Religion: Remaining Human in a Co-created World*, (Templeton Science & Religion, 2009), 69.

⁵⁴ Tillich, P. *Systematic Theology Volume 1*, (Chicago: Chicago University Press, 1951), 258-60.

⁵⁵ Herzfeld, *In Our Image*, 18.

⁵⁶ Cortez, *Theological Anthropology*, 19.

debate in both Christian anthropology and AI research is between the functional and the relational models, with each approach locating ‘the core of our humanity in a different sphere and suggests different trajectories for the project of AI.’⁵⁷

The Functional Interpretation

A shift from the substantive approach to something new began with the work of theologians such as Johannes Hehn and Gerhard von Rad, where the *imago Dei* was located in human activity or agency in the created world. Alistair McFadyen locates the key difference between the functional and substantive models in their respective treatment of the image as verb rather than noun⁵⁸, a useful distinction. Functional approaches to the motif take scripture such as Genesis 1:26 and apply it to how humans interact and exercise dominion over the rest of creation. For Herzfeld, this is the key to this interpretation, with humanity ‘acting as God’s deputy on earth.’⁵⁹ This model raises questions around where the boundaries of our dominion lie, are we limited or restricted in how far we can go? Is it possible for humans to overextend their mandate and threaten their relationship with the image? Equally, what are the consequences if we do not participate enough as agents of God’s creation, could there be a command found within the functional model for humans to engage and create with technology to fulfill their remit?

By the 1980s the functional view of the *imago Dei* was found in AI and robotics as researchers aimed to create technology that could perform human tasks. It is worth considering at this point whether the ability to conduct these tasks indicates the possession of agency, the ability to act with intent to fulfill a need or obtain a desire, or not. An entity must act with intention to be in ownership of agency, and while contemporary AI can calculate, assess options, and make choices in a way that seems intentional, I do not believe that it currently possess agency, a view supported by Swanepoel in her 2021 study of this issue.⁶⁰ When considering the AI

⁵⁷ Herzfeld, *Creating in Our Own Image*, 312.

⁵⁸ McFadyen, Alistair. (2012) *Imaging God: A Theological Answer to the Anthropological Question*, *Zygon*, 47:4, 919.

⁵⁹ Herzfeld, *Creating in Our Own Image*, 306.

⁶⁰ Swanepoel, D. ‘Does Artificial Intelligence Have Agency?’ in R. Clowes, K. Gärtner and I. Hipólito (eds.) *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts*, (Springer, 2021), 102.

depicted by Ishiguro in *Klara and the Sun* in Chapter Four I will return to this question of agency again.

The roboticist and computer scientist Hans Moravec took a functional reading of this situation, stating in 1988 that computers will eventually displace humans from essential jobs and roles.⁶¹ Technology has had notable success in this area, with humans made redundant from a number of positions and advancements in processing power leading computers to be able to perform increasingly complex tasks. The computer Deep Blue achieved victory over Garry Kasparov in 1997 and in 2011 IBM built on this success developing Watson, a question answering computer system that analysed natural language questions and content to a high enough standard to win against champion Jeopardy! players, using ‘computational linguistics, information retrieval, knowledge representation and reasoning, and machine learning’⁶² to do so.

According to Moravec’s Paradox, it is relatively easy to train computers to do things that most humans find difficult, such as mathematics and logic, but it is very hard to train them to do the things humans usually find easy, such as walking and image recognition, a shortcoming that inevitably limits their current scope. Subsequent application of Watson in clinical research has demonstrated it can achieve a type of learning through its algorithms, which recognise patterns and behaviours to create its own logic. But these machines behave differently to humans in that they cannot think out with the perimeters of what they have explicitly been programmed to do and they cannot determine the whys or the causes of what they calculate. A further issue within the functional model is that within this interpretation of the image, what do we define as AI? What separates AI from an everyday computer application? The functional model’s focus on task and action means that virtually every programme that can accomplish a purpose would fall into the AI category and, as Herzfeld highlights, ‘it would be ludicrous to consider all programs to be artificially intelligent.’⁶³

⁶¹ See Moravec, H. P. *Mind Children : the Future of Robot and Human Intelligence*, (Cambridge: Harvard University Press, 1988).

⁶² IBM. (2012). This is Watson, *IBM Journal of Research and Development*, Issue 3.4: May-June 2012, [Online] <https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6177717> (Accessed 01 August 2022).

⁶³ Herzfeld, *Creating in Our Own Image*, 308.

I agree with Herzfeld's conclusion that the functional approach creates an 'unsatisfying image of humankind and an equally unsatisfying image of God.'⁶⁴ The model raises the concerns expressed by Haraway in Chapter One of anthropocentrism and there is also potential for the exploitation of nature to occur. From an AI and robotics point-of-view it can also create a fear around human replacement not found in the substantive model, a theme explored by Ishiguro in his novels *Klara and the Sun* and *Never Let Me Go* that will be discussed further later. As in the substantive interpretation, the functional approach does not consider relationships as a valuable part of the model, which brings us to the third view of the *imago Dei*: the relational.

The Relational Interpretation

The relational view is the third primary model of *imago Dei*, in which the image and likeness of God is found in human-divine and human-human relationships. It is not seen as part of the human structure, and it is not 'accomplished' as it is in the functional example, but instead God implores humans to relate to them, and to each other, and the image is then present in those humans. Herzfeld describes the relational model as presupposing 'no essential difference between humans and other creatures, just as it does not posit an essential similarity between humans and God. Instead, it focuses on the calling of human beings to be in relationship'⁶⁵ In observing the *imago Dei* within the relational interpretation, Herzfeld examines in depth the work of Karl Barth, viewing him as a key proponent of the relationship as image.⁶⁶ For Barth, the image of God is not substantive as it does not exist in what human beings are, it is not functional as it is not concerned with what humans do, but exists in the relational, with the human being as a counterpoint to God. His argument finds its roots in Genesis 1:26-27, in the two key phrases 'let us make humankind in our image' and 'male and female he created them'.

In *Church Dogmatics* Barth writes:

If the divinity of the man Jesus is to be described comprehensively in the statement that He is man for God. His humanity can and must be described no less

⁶⁴ Ibid, 313.

⁶⁵ Herzfeld, *In Our Image*, 90.

⁶⁶ Herzfeld, *Creating in Our Own Image*, 308.

succinctly in the proposition that He is man for man, for other man, His fellow.⁶⁷

This double proclamation of Jesus's existence gives the form content of humanity as the I-Thou relationship, consisting of three different but related concepts; the I-Thou relationship within God Himself, the I-Thou relationship between God and human being and the I-Thou relationship between human beings. This assertion merges the *imago Dei* narratives of the Old and New Testaments, arguing that Adam and Eve were created in a harmonious relationship with God and all of creation, but these relationships were damaged by the fall and only in Christ is the image restored. Much of Barth's writing focuses on God's work within the created world, the incarnation of Christ being the pinnacle of this action. He states that 'all things in heaven and earth are the objects of the divine purpose. But this purpose is not disclosed in all things; it is disclosed only in man. The cosmos surrounding man is not alien to God.'⁶⁸ For Barth, the image is not a quality, humans do not individually hold it, but instead it exists in a relationship to God and in our relationships with each other. A criticism of Barth's relational interpretation is found in his neglect of the general exegetical agreement surrounding the plural 'let us' as referring to heavenly court rather than a multiplicity within the Godhead.⁶⁹ While Herzfeld is in agreement with Barth regarding the importance of relationality in interpreting the *imago Dei*, she notes that theologians are divided in their interpretation of Barth, with disagreement concerning the 'details of what constitutes authentic relationship'⁷⁰ and the dominance of male-female relationships in his theology.

Critics of the relational view cite difficulties in coherently defining what the image actually is, the theologian Millard Erickson highlights that this model raises the question do all humankind inherently possess the image, or is it only experienced when they establish relationships with God and each other?⁷¹ Gerald Bray rejects Barth's claim that the image can only be realized in community as this leaves the

⁶⁷ Barth, K. *Church Dogmatics: III/2*, trans. J.W. Edwards, O. Bussey and Harold Knight, (Edinburgh: T&T Clark: 1958), 135.

⁶⁸ Barth, *Church Dogmatics: III/2*, 16.

⁶⁹ Von Rad, Gerhard, *Genesis: A Commentary*, (Presbyterian Publishing Corporation, 1973), 57-58.

⁷⁰ Herzfeld, *In Our Image*, 25.

⁷¹ Erickson, M. J. *Christian Theology*, (Baker Books, 1983), 508.

solitary individual as ‘an incomplete image and therefore an incomplete person.’⁷² Bray acknowledges that God states that man should not be alone⁷³, but finds no indication that the *imago Dei* would be imperfect in those without companionship. Instead, Bray finds the image ‘is designed to establish a relationship of community or fellowship primarily between man and God, not between male and female or between man and the lower creation.’⁷⁴

Herzfeld finds the most potential to understand humankind in a world with robots and AI in the relational interpretation due to its emphasis on human connection with both God and other human beings, inclusive of all people. She comments that the ‘image of God in humankind has become one cornerstone of Christian anthropology, a locus for understanding who we are in relation to both God and to the world’⁷⁵ and her focus on Priestly biblical writing⁷⁶ demonstrates how deep the roots of theological discussion concerning robots and AI can be viewed. As a model the relational also has the potential to dispel the concerns examined under the functional interpretation around replacement, because human dignity is not inextricably linked to the completion of tasks here. Herzfeld writes that if our focus is our relationships, then we do not need to fear replacement⁷⁷, although the potential for robots and AI to replace us would certainly affect the relationship.

Alan Turing’s 1950 essay *Computing Machinery and Intelligence* concerned the nature of machine intelligence rather than the computational power itself and anticipates possible objections or concerns around machine intelligence. In response to the theological argument that no animal or machine can think because they have not been granted an immortal soul, Turing replies that:

In attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates.⁷⁸

⁷² Bray, Gerald. (1991), The Significance of God’s Image in Man, *Tyndale Bulletin* 42:2, 223.

⁷³ In Genesis 2:18.

⁷⁴ Bray, *The Significance of God’s Image in Man*, 223.

⁷⁵ Herzfeld, *In Our Image*, 6.

⁷⁶ Referring to the books of the Bible that make up the Pentateuch, Genesis, Exodus, Leviticus, Numbers and Deuteronomy.

⁷⁷ Herzfeld, *Creating in Our Own Image*, 312.

⁷⁸ Turing, A. (1950), Computing Machinery and Intelligence, *Mind*, 59:236, 443.

He also observes that the concept of the soul is different in different religions and because it is not a universal constant, we cannot use it as leverage in the argument against AI. He continues that theological objection to AI limits the omnipotence of God, writing that ‘should we not believe that He has freedom to confer a soul on an elephant if He sees fit?’⁷⁹

Herzfeld locates the relational view of the *imago Dei* in modern AI research in the commitment to social robotics. She describes the Turing Test as fitting into the relational model because of its focus on the execution of language and conversational abilities, relating to a human being in dialogue, rather than just performing a set task. She argues that the test determines intelligence using relationality because ‘discourse is unique among human activities in that it subsumes all other activities within itself’⁸⁰. Herzfeld finds support for this relational interpretation in the value the Turing Test places on mirroring humans, including making errors and perhaps demonstrating inconsistency, writing that the ‘mistakes of hesitancy....are hallmarks of human functioning.’⁸¹ However, there are difficulties in accepting the Turing test as relational. The technology that currently takes part is disembodied intelligence, there can be no relationship between it and a physical body or its environment. Herzfeld herself recognises this fact in the work of Barth, remarking that he has a ‘highly embodied view of what constitutes authentic relationship’⁸² but fails to highlight the Turing Test has nothing at all to do with embodied relationality. She continues that for Barth, relationality includes ‘the ability to look the other in the eye, to speak and hear, and to give aid,’⁸³ also all missing from the test. Herzfeld moves into more relational territory with her discussion of Cog and Kismet, robotic experiments in affective technology designed to recognise and identify emotions. They have the ability to learn, make eye contact and hold objects and are described by Herzfeld as ‘a co-respondent with which to relate.’⁸⁴ While Cog and Kismet take us further into the relational model

⁷⁹ Ibid.

⁸⁰ Herzfeld, *A New Member of the Family?*, 242.

⁸¹ Herzfeld, *In Our Image*, 46.

⁸² Herzfeld, Noreen. (2005). Terminator or Super Mario: Human/Computer Hybrids, Actual and Virtual, *Dialog* 44:4, 351.

⁸³ Ibid.

⁸⁴ Herzfeld, *In Our Image*, 82.

than the Turing Test, they still fail to reference the ability to display many human qualities, have no familial relationships, and are not part of wider society.

One question Herzfeld's work raises concerns feasibility; the emergence of AGI, or strong-AI, would be a requirement of her proposition. AGI is the representation of generalised human cognitive abilities in software so that if confronted with an unfamiliar task, the system could work towards finding a solution. While theoretical at present, it is a form of AI that would be able to replicate human functions such as reasoning, problem-solving and planning. Herzfeld writes confusingly about this prospect at times, on the one hand she seems to propose an optimism about the possibility that future AI could bring, but on the other she comments that no significant progress has been made in AI research since its inception in the 1950s.⁸⁵ Marius Dorobantu is in agreement with Herzfeld concerning the stilts in progress that can be found, referring to them as the 'AI winters.'⁸⁶ However, in 2014 a survey of expert opinion was conducted concerning the probability of strong-AI existing before 2050, with 50 percent of responding AI experts reporting they thought it possible.⁸⁷ In *Klara and the Sun* machine intelligence has reached the point where the creation of capacities once thought to be uniquely human, such as learning, creativity and even intuition, have become possible with the development of Artificial Friends, raising important questions in Ishiguro's created world around human distinctiveness.

In many respects, Herzfeld's support for the relational model is persuasive. Her approach to robotics and AI in context of the *imago Dei* has given me a far greater understanding of the theological and ethical questions surrounding this topic. The balance between disciplines in her interdisciplinary approach is fruitful and reiterates to me the importance of understanding the benefits and downsides of both the literary and theological in the methodological approach I am proposing.

⁸⁵ Ibid, x.

⁸⁶ Dorobantu, M. (2019), Recent Advances in Artificial Intelligence (AI) and Some of the Issues in the Theology and AI Dialogue, *ESSSAT News & Reviews*, 29:2, 6.

⁸⁷ Muller, V. and Bostrom, N. Future Progress in Artificial Intelligence: A Survey of Expert Opinion in *Fundamental Issues of Artificial Intelligence*, edited by Vincent Muller, (Berlin: Springer, 2014), 555.

Created Co-Creators?

Herzfeld suggests we need to look again at the *imago Dei* and examine not only the created human, but also the creative potential she believes is a key element of being human. Herzfeld considers our responsibility in relation to technology by understanding humans as ‘co-creators’, examining where the balance between ‘co’ and ‘creator’ exists and concluding we should focus on the former, believing that the aim of AI is to create an ‘other’ in our own image.⁸⁸ There is some historical support for Herzfeld’s position, such as Paul Tillich’s proposal that human life consisted of three integral functions; self-integration, self-transcendence and self-creation. The latter addresses the creation of culture and Tillich recognises language and technology as a key concept for this dimension of the human life.⁸⁹

The Lutheran theologian Philip Hefner is a more recent proponent of the idea of the self-creation of culture, including technology, and views it as being essential to humankind. He first used the term ‘created co-creator’ in 1984 and went on to evolve it in later works, eventually forming a theory examining God’s purpose for creation. He asserts that the doctrine of creation is Christianity’s affirmation of a relationship between God and the rest of creation, which when combined with his strong emphasis on the importance of belonging, leans Hefner’s theory towards the relational interpretation, although he can also sit comfortably within the functional view. For Hefner, the purpose of human beings as created co-creators is to be ‘the agency, acting in freedom, to birth the future that is most wholesome for the nature that has birthed us.’⁹⁰ Humans are not equal with God, but we do have agency to work for Them within creation, the playing out of another creator and co-creator relationship. This raises questions that will be examined later in the context of the AI in Ishiguro’s novel, including could AI supersede humans? If what makes us human is not distinct enough from what makes a machine a machine, if there is little or no discernible difference, does this leave humans in a precarious

⁸⁸ Herzfeld, N. ‘Co-Creator or Co-creator? The Problem with Artificial Intelligence’ in U. Görman (ed.) *Creative Creatures: Values and Ethical Issues in Theology, Science and Technology*, (Bloomsbury Publishing, 2005), 46.

⁸⁹ Tillich, P. *Systematic Theology*, Vol. 3, (Welwyn: James Nisbet & Co, 1964), 32-34 and 61-66.

⁹⁰ Hefner, P. *The Human Factor: Evolution, Culture and Religion*, (Minneapolis: Fortress Press, 1993), 27.

position? If we take Herzfeld's idea that a goal of AI is to create an 'other' in our image, what part of the human do we wish to pass on? The choices we make in answering these questions have implications for both our self-understanding and our future co-existence with what we create.⁹¹

Hefner finds an optimism in this potential future, asserting that 'to be created in the image of God implies that humans can be the vehicle for grace toward the creation, in a way that is somehow reminiscent of God's graciousness.'⁹² In his exploration of cyborg theology, a hopefulness is also proposed by Scott Midson, who states that when we consider where the boundaries and limits to our creation are, he finds that humans are unique in that they can create and keep at least partial control over *what* they create, a concept he calls 'capacity'.⁹³

Again, in this creative dynamic the human is not compared to God, but to human and robot as created other. While Herzfeld clearly sees parallels between the creative element present in the *imago Dei* and activity within AI research, they are not exact parallels surmising that:

If we hope to find in AI that other with whom we can share our being and our responsibilities, then we will have created a stand-in for God in our own image...bound to be a disappointment.⁹⁴

Allen Emerson and Cheryl Forbes are in agreement with this view, considering robots as new Adams, with whom 'we will stand to this creation as God stands to us'.⁹⁵

Eschatological *Imago Dei*; A Fourth Way?

The eschatological interpretation of the *imago Dei* is proposed by a number of theologians, most notably Wolfhart Pannenberg and Jürgen Moltmann. Essential to Pannenberg's anthropology is Johann Gottfried Herder's idea of the image, which considers all human reality and the course of human history as a process of education to humankind. For both theologians the image of God is present in

⁹¹ Herzfeld, *Creating in Our Own Image*, 304.

⁹² Hefner, *The Human Factor*, 238.

⁹³ Midson, S. *Cyborg Theology: Humans, Technology and God*, (I.B. Tauris, 2017), 54.

⁹⁴ *Ibid.* 313.

⁹⁵ Emerson, A. and Forbes, C. (1984). Living in a World with Thinking Machines: Intelligence Will Not Separate People from Machines, *Christianity Today* 28:2, 14.

silhouette form in humans by the virtue of creation. Pannenberg relates the full image to exocentricity, an openness to both the world and the possibility of relationships with others, including with God, that brings transcendence to human experience. Pannenberg locates Herder's anthropology in a Christological basis and his notion of the *imago Dei* is realised through the destiny of humanity to find the purpose of their existence out with themselves, towards a future destiny.⁹⁶ F. LeRon Shultz argues the eschatological is not a new category of the image at all, but rather one that includes a combination of factors from the substantial, functional and relational models detailed above.⁹⁷ Shultz's line of reason fails to fully recognise the uniqueness of positions such as Pannenberg's, where the image is placed in relation to eschatology and salvation as well as creation. The eschatological model will be an interesting one to explore in this dissertation as its content relates well to themes presented by Ishiguro.

While my analysis has classified interpretations of the *imago Dei* into four main categories, there are elements of overlap within the models. Cortez proposes that from a standalone position, neither the substantive, functional or relational model is satisfactory, and instead he proposes a 'multifaceted'⁹⁸ approach which unites them into one system. It does seem unlikely that any of the methods could operate in complete isolation and it may be argued that each historical model builds on the previous, coming together as the pieces that make up the image. All three models have implications for where we see ourselves in relation to the created world.

Herzfeld finds many similarities between how researchers in the AI field understand intelligence and how theologians view what it means to be made in the image of God. For her, AI can identify with the substantive mode in its pursuit for cybernetic immortality⁹⁹ and she quotes Moravec and his 'mind-children'¹⁰⁰ who will outlive us. AI can also operate in the functional model with its ability to support human dominion; it can bolster the information and knowledge we have available to us allowing us to extend our influence. The relational view can be found in the relationships that will develop between humans and AI, although she

⁹⁶ Pannenberg, W. *Anthropologies in Theological Perspective*, (T&T Clark, 1985), 43-79.

⁹⁷ Summarised in van Huyssteen, *Alone in the World?*, 139-143.

⁹⁸ Cortez, *Theological Anthropology*, 18-21.

⁹⁹ Herzfeld, *In Our Image*, 69.

¹⁰⁰ See: Moravec, *Mind Children*.

warns against believing we can create artificial companions who will substitute for our relationships with God and other humans. Herzfeld concludes that ‘the real debate, both in theology and AI, is now between the functional and relational camps’.¹⁰¹

The development of AGI would be a major setback to the substantive model. If technology could outperform human beings at all tasks, where does that leave human uniqueness and distinctiveness in this interpretation? The emergence of strong-AI would also have implications for the functional and relational models of the image. In the former, God has elected humans to act within creation and strong-AI would be able to work in the world with at least the same ability as humans. The relational model places the image in the human capacity and inclination for relationships with both God and each other. But is the ability to engage in and conduct relationships all that is needed in the relational interpretation? Within Barth’s relational model there is a suggestion that personhood and agency are required elements for meaningful relationships. For Barth, humankind is made in the *imago Dei* because they are the ‘Thou’ God addresses and who can answer back, but is this a dimension accessible to AI? In Chapters Three and Four I turn to literature and, as outlined in Chapter One, use Ishiguro’s novel *Klara and the Sun* to build a thought-experiment to see what this speculative novel has to say about how strong-AI may or may not be accommodated in any of the theological models of the *imago Dei* discussed here.

¹⁰¹ Herzfeld, *A New Member of the Family?*, 243.

Chapter Three: AI and Literature

‘One of mankind’s most cherished dreams - in religious, scientific, and artistic circles - has been the creation of humanoid life’¹⁰² writes Robert Geraci, professor of religious studies. In the modern world, AI and robotics research hold the most promise for making this a reality, with the cyberneticist Kevin Warwick defining AI as:

Not the study of computers, but of intelligence in thought and action. Computers are often its tools, because its theories are usually expressed as computer procedures that enable machines to do things that would require intelligence if done by people.¹⁰³

Warwick’s definition is interesting as it represents an openness to the inclusion of all disciplines with an interest in intelligence, rather than restriction to the fields of computer science and electrical engineering found in other definitions.¹⁰⁴ He also makes the important connection between the need to understand human intelligence to be able to understand artificial intelligence.

The cognitive scientist Margaret Boden identifies two main aims in the development of AI; the technological, used to get useful things done, and the scientific, where it is used to ‘help answer questions about human beings and other living things.’¹⁰⁵ She identifies five key types of AI: the classical or symbolic, artificial neural networks or connectionism, evolutionary programming, cellular automata and dynamical systems,¹⁰⁶ with the possibility of creating virtual machines that are hybrids of two or more of the categories. There are many excellent sources that discuss the various meanings of the term AI and detail its

¹⁰² Geraci, R. (2008). Apocalyptic AI: Religion and the Promise of Artificial Intelligence, *Journal of the American Academy of Religion*, (76:1), 138.

¹⁰³ Warwick, K. *In the Mind of the Machine: The Breakthrough in Artificial Intelligence* (London: Arrow, 1998), 94.

¹⁰⁴ For example, see Anne Foerst’s definition in: Foerst, *God in the Machine*, 66.

¹⁰⁵ Boden, M. *Artificial Intelligence: A Very Short Introduction* (Oxford: Oxford University Press, 2018), 2.

¹⁰⁶ *Ibid*, 5.

history and projected future¹⁰⁷ and only the most cursory examination of the field can take place here.

A Brief History of AI

The origins of modern AI can be found in the automata of the past, the word itself coming via Latin from the Greek *automatos* or ‘acting of itself’. One of the oldest known examples was the clock built in Cathédrale Notre-Dame in Strasbourg during the 1350s, with a centrepiece of a mechanical rooster that could spread its feathers and crow. The clock also featured the three biblical kings who would appear and bow before Mary and Jesus. In the seventeenth century the Strasbourg clock received a lot of interest from natural philosophers, including René Descartes and Robert Boyle, who saw analogues to universal order in the clock. Descartes’s interest in automata was reputedly linked to his theory that living things were biological machines that functioned in a similar manner to clockwork.¹⁰⁸ Other early automatons included *karakuri* in Japan, traditional mechanised puppets or automata that became popular entertainment from the seventeenth century. The period 1860 to 1910 is known as the golden age of automata, with French companies in particular exporting clockwork automata and mechanical singing birds worldwide.

By 1936 Alan Turing had demonstrated the possibility of a single machine that could be used to calculate any computable sequence, a system that would eventually become known as a Universal Turing Machine. His belief that AI was a meaningful possibility was strengthened in the 1940s by the work of scientists such as the neurologist Warren McCulloch and the mathematician Walter Pitts, who brought Turing’s work into dialogue with Charles Sherrington’s neural synapses theory and the philosopher Bertrand Russell’s propositional logic.¹⁰⁹ The first

¹⁰⁷ For example, see: Wooldridge, M. *The Road to Conscious Machines: The Story of AI* (Pelican, 2020), Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence* (London: Penguin Books, 2018) and Mitchell, M. *Artificial Intelligence: A Guide for Thinking Humans* (Pelican, 2020). Texts that deal in detail with AI and theology and/or religion include: Mercer, C.R. & Trothen, T.J. *Religion and the Technological Future: An Introduction to Biohacking, Artificial Intelligence* (Palgrave Macmillan, 2021) and Midson, S. *Cyborg Theology: Humans, Technology and God* (Bloomsbury Publishing, 2017).

¹⁰⁸ Wood, G. *Living Dolls: A Magical History of the Quest for Mechanical Life* (London: Faber and Faber, 2002), 5.

¹⁰⁹ See McCulloch, W.S. and Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity, *The Bulletin of Mathematical Biophysics*, (5), pp.115-133.

academic conference on AI was held in the United States in 1956 and Marvin Minsky co-founded the Artificial Intelligence Laboratory at Massachusetts Institute for Technology (MIT) two years later. By the mid-1980s, Emerson and Forbes proposed that the evolution of robots and AI might be more revolutionary than even the invention of writing and printing,¹¹⁰ and while we are yet to reach the heights of their predictions, AI has become increasingly intertwined in our daily lives. As well as its practical applications in our homes, vehicles and workplaces, some AI has already left Earth and can be found in robots sent to Mars or in satellites orbiting space. This interweaving of AI into human life is clearly an ethical and theological challenge theologians should respond to, continuing the biblical tradition of speaking about the moral issues of the day. Theological responses have the potential to offer alternate points of view to put forward from secular sources and these contributions may be contain insights of value to wider society.

If we look to the future, the realisation of AGI would be a major step beyond current developments. When we imagine systems performing equally to a human, we move from questions of science to ones traditionally regarded as philosophy; will they have real intelligence? Do they possess real understanding? Do they have free will and free choice? Are they conscious? For many people, the instinctive answers to these questions may be a straightforward 'no', but we cannot answer either way for sure. As strong-AI is not yet a reality we can only speculate on how it may be achieved, what it may look like and how it will affect society, culture and relationships. Literature functions as a privileged space in which these explorations can be made.

¹¹⁰ Emerson and Forbes, *Living in a World with Thinking Machines*, 14.

A Brief History of AI in Literature¹¹¹

Adrienne Mayor, folklorist and historian of ancient science, writes that Greek myth contains many examples of those who were ‘made’ rather than born. John Cohen, in *Human Robots in Myth and Science*, agrees, finding the ‘first ancestors of modern automata in the twilight figures of remote mythology.’¹¹² Looking at the stories of Jason and the Argonauts we find the tale of the bronze robot Talos, first mentioned by Hesiod in around 700BC, a giant metal man built by Hephaestus, the Greek God of invention and blacksmithing. At his core Talos had a channel carrying a life source the Greeks called *ichor*, powering him to march around the island of Crete three times a day throwing boulders at enemy ships. As is often the case with representations of technology in literature, Talos is eventually destroyed, when Medea exploits a weakness in his ankle. Hesiod was also the first known writer to tell the myth of Pandora, his version in *Theogony* describes her as a mythical artificial being built by Hephaestus and sent to Earth by Zeus as punishment for humans acquiring fire, this contrasts with the more sanitised later retellings of the story. In East India in the fifth century BC there were legends that King Ajatashatru had the Buddha’s tomb guarded by giant robot guardians, the bhuta vahana yanta, or ‘spirit movement machines’ in Sanskrit. In Jewish folklore golems were anthropomorphic beings usually created from clay or mud that would only become animated when their human masters placed a slip of paper in their mouths or foreheads with the name ‘God’ written on it. Mayor writes that these myths ‘represent the earliest expressions of the timeless impulse to create artificial life’¹¹³ and a certain futuristic anticipation lies in these stories.

¹¹¹ For a more in-depth exploration of AI in literature several notable sources are available. D. Porush’s *The Soft Machine: Cybernetic Fiction* (London: Routledge, 2018) is a study of cybernetics and AI in postmodern literature that examines the work of Pynchon, Beckett, Burroughs and Vonnegut¹¹¹. S. Vint and S. Buran edit *Technologies of Feminist Speculative Fiction: Gender, Artificial Life and the Politics of Reproduction*, (Palgrave Macmillan, 2022). a collection of essays concerning feminist speculative fiction and AI,¹¹¹ and B. Dainton, W. Slocombe and A. Tanyi edit *Minding the Future: Artificial Intelligence, Philosophical Visions and Science Fiction* (Springer, 2021) a volume of work which discusses the philosophical aspects in connection with AI, making use of speculative and SF literature such as Asimov, Leckie and I.M, Banks to do so.

¹¹² Cohen, J. *Human Robots in Myth and Science* (London: George Allen & Unwin, 1966), 15.

¹¹³ Mayor, A. *God and Robots: Myths, Machines, and Ancient Dreams of Technology* (Princeton University Press, 2018), 1.

Moving forward to the nineteenth century, we begin to find in literature the type of robot familiar to us today, created intentionally by humans with their construction emphasising science and engineering rather than magic and divinity. Examples from this period include the 1868 novel *The Steam Man of the Prairies* by Edward S. Ellis, the frontier adventure features a bipedal steam-powered android who goes on adventures with his teenage creator. In 1920 Karel Čapek's play *RUR (Rossum's Universal Robots)* debuted at the National Theatre in Prague, featuring an inventor who manufactures a race of artificial workers made of living tissue and flesh that resemble more what we would now call androids or clones rather than metal men. The robots Čapek creates are so lifelike they can be mistaken for humans, they can think for themselves and can also feel pain, a feature their inventor purposely included in their programming for commercial reasons as a guarantee the robots will take care of their physical bodies. By the play's conclusion the robots have revolted and killed all but one person, with the inevitable demise of human life the audience is left with two robots who have fallen in love, framed by Čapek as a new iteration of Adam and Eve. Robots and androids have been used by writers and storytellers since *RUR's* debut to examine two key questions; what would society look like if humans were liberated from work? And could these inventions develop agendas of their own, perhaps in conflict with the interests of humankind?

In discussing AI in film, Herzfeld identifies two general categories; the cautionary tale and the wish fulfilment tale.¹¹⁴ This division is mirrored in literature too with SF and speculative fiction having a long history of portraying the dichotomy of 'good' and 'bad' technology, from Čapek's killer robots to Lester del Rey's *Helen O'Loy*, a household robot who manages to overcome her programming to develop emotions and fall in love with a human. In 1950 Isaac Asimov addressed this issue in his 1942 short story *Runaround*, published in the collection of short stories *I, Robot*, in which he reveals his Three Laws of Robotics. Asimov chose to write about robots and AI in response to what he saw as the 'unrealistically wicked or unrealistically noble'¹¹⁵ accounts found in SF to that point. The Laws were designed to promote human-robot relations and it is worth outlining them in their

¹¹⁴ Herzfeld, *In Our Image*, 55.

¹¹⁵ Isaac Asimov quoted in: Jordan, J. M. *Robots* (Cambridge, MA: MIT Press, 2016), 32.

entirety as they will be relevant to the examination of *Klara and the Sun* in the next chapter:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.¹¹⁶

Asimov gives ethical instructions that have a lot in common with the functional view of technology discussed in the previous chapter. They are built on the *difference* between humans and robots, finding the latter very much subservient to the former. In this view humans have a uniqueness both in their ability to create robots and also in their control of them.

Within contemporary speculative fiction there were several recent literary works where the theological motif of the *imago Dei* could have been considered in context with AI and machine consciousness, including Ian McEwan's 2019 novel *Machines Like Me*. This story is set in a speculative alternative reality where Alan Turing did not die in 1954, but instead lived to outline the theoretical principles allowing for the creation of humanlike robots by the early 1980s. Turing himself appears as a character in the story, functioning almost as the novel's conscience. The trajectory of the text follows Adam, a strong-AI robot configured so well that some cannot tell him apart from a human, and his life and relationship with Charlie, his owner, and Charlie's girlfriend Miranda, who at the novel's opening he co-authors Adam's personality with. Jeanette Winterson's *Frankissstein* weaves an exploration of AI and transhumanism into a love story, re-telling Mary Shelley's classic story with Victor Stein who intends to resurrect the dead using AI. Other contemporary novels that could have been considered include David Mitchell's *Cloud Atlas*, Becky Chambers's *A Closed and Common Orbit* and *Exegesis* by Astro Teller.

In all these works stories about artificially created life often come back to address similar questions to those asked in the previous chapter on the *imago Dei*; what does it mean to be human? Why do humans create? Are there, or should there be,

¹¹⁶ Asimov, I. Runaround in *I, Robot* (New York City: Doubleday, 1950), 40.

limits on human creativity? The monster says to Victor in Shelley's *Frankenstein* 'I ought to be thy Adam, but I am rather the fallen angel'¹¹⁷ but what should the nature of human relationships and their creations be? To explore these issues in greater depth I now turn to examine Ishiguro and his novel *Klara and the Sun*.

Kazuo Ishiguro and *Klara and the Sun*

Kazuo Ishiguro was born in Nagasaki, Japan in 1954 and moved to Guildford, Surrey with his parents when he was five. He is a celebrated author of eight novels and was awarded the 2017 Nobel Prize in Literature. The work of contemporary speculative fiction chosen for consideration in this dissertation is his 2021 novel *Klara and the Sun*. It was selected because it is a very recent publication that has not yet received a huge amount of criticism or academic study; there were fewer than thirty academic journal articles available on the University of Glasgow's library search at the time of writing in July 2022. The story is told from the point of view of Klara, an AI, which interested me as this is not just a novel *about* machine intelligence but narrated by one. This is a relatively unusual concept in literature, with the notable exceptions of *Saturn's Children* by Charles Stross, where all the characters are AI, and Asimov's novelette *The Bicentennial Man*. The epistolary narrative in Shelley's *Frankenstein* allows for a portion of the text to be told in the direct voice of Frankenstein's creation so could also be included in this category.

At first glance *Klara and the Sun* is not necessarily a novel that has much to say about religion or theology and Ishiguro is not necessarily an obvious choice of author for an examination of a topic such as the *imago Dei*. He revealed in an interview that he is wary of labelling himself as an atheist, commenting that 'religions are things that we make up. It's another question whether God is made up, but surely religions are man-made things.'¹¹⁸ He talks of the lack of religion in his childhood, not knowing until he was fifteen that the family's official religion

¹¹⁷ Shelley, M. *Frankenstein; Or, The Modern Prometheus*, (Fine Communications, 2003), 98.

¹¹⁸ Delistraty, C. (2015). Lost Toys and Flying Machines: A Talk with Kazuo Ishiguro in *The New Yorker* [Online] Available at: <https://www.newyorker.com/books/page-turner/lost-toys-and-flying-machines-a-talk-with-kazuo-ishiguro#:~:text=Ishiguro's%20views%20on%20religion%20are,things%20that%20we%20make%20up>

was Buddhist Zen.¹¹⁹ The absence of religion is a noticeable characteristic of Ishiguro's novels, except for the destructive role the early medieval church plays in agreeing to forget things in *The Buried Giant*. Whilst Ishiguro's novels do not implicitly comment much on religion, they are deeply moral stories that ask key questions about the human condition, his narratives often suggesting something significantly deeper beneath the surface. Issues with a theological angle are raised in many of Ishiguro's more recent novels, such as the matter of whether the human clones in *Never Let Me Go* have souls and the exploration of faith versus rationality and reason in *Klara and the Sun*. As Margaret Atwood points out, it is also worth remembering that an Ishiguro novel 'is never about what it pretends to be about.'¹²⁰

Some authors writing within the speculative genre have highlighted that their novels were not intended to be read the way some have interpreted. For example, Atwood describes the theocratic regime in *The Handmaid's Tale* as 'not genuinely Christian'¹²¹ although theologians such as Carol Christ continue to read it as such. In an analysis of José Saramago's novel *Blindness*, Andrew Hass proposes that awareness of the author's politics and atheism may automatically suggest reading the work through these lenses, a view both he and I disagree with. I would subscribe to Roland Barthes's theory in *The Death of the Author* that the reader as producer is free to find their own meaning in a literary work. Additionally, while our society may be increasingly secularized, this does not mean that theological ideas and symbols do not have their role in contemporary lives and beliefs. As Scott Midson informs us, 'theology...tells particularly pervasive stories that shape our attitudes and understanding, even without us perhaps realizing so at a more manifest level.'¹²²

Summary of *Klara and the Sun*

¹¹⁹ Gross, T. (2021). *Kazuo Ishiguro Draws on his Songwriting Past to Write Novels About the Future* [Online] Available at: <https://www.npr.org/transcripts/978138547> (Accessed 01 August 2022).

¹²⁰ Margaret Atwood quoted in: Thomas, P.L. *Science Fiction and Speculative Fiction: Challenging Genres* (Sense Publishers, 2013), 168.

¹²¹ Williams, L. (2017). *Margaret Atwood on Christianity, 'The Handmaid's Tale', and what Faithful Activism Looks Like Today* [Online] Available at: <https://sojo.net/articles/margaret-atwood-christianity-handmaid-s-tale-and-what-faithful-activism-looks-today> (Accessed 01 August 2022).

¹²² Midson, *Cyborg Theology*, 28.

Klara and the Sun was Ishiguro's first novel since winning the Nobel Prize in literature in 2017. It is set in a speculative near future that feels somewhat like our present in which AI has taken on many of the roles previously performed by humans, although there is no backstory explaining how the leap in technology to create strong-AI such as Klara occurred. Factions are growing in the cities and the impression of a society on the edge of collapse is suggested in the fleeting references to the wider world Ishiguro grants us.

The novel's eponymous main character is Klara, an 'Artificial Friend' or 'AF', a solar-powered humanoid robot. She opens the novel in a department store awaiting purchase by a parent to become the companion of a child. Each AF in the shop is slightly different, and we find out from the manager that Klara's special talent is her perceptiveness with feelings, describing her as having an 'appetite for observing and learning ... [and] the most sophisticated understanding of any AF in this store.'¹²³ While sitting in the prized window seat at the store, a girl named Josie is drawn to Klara and promises to return and purchase her. Eventually she does come back with Chrissie, her mother, and purchases the AF.

Klara moves into the family home with Josie, Chrissie and the housekeeper Melania. We find out early in the novel that parents may choose to genetically modify their children in a process called 'lifting'. This practice is never fully explained to the reader but we do find out it has created a 'superclass' of adolescents who are privileged with university admissions and employment opportunities not usually conferred on the 'non-lifted'. Lifting can occasionally be a fatal process, and we discover that Josie's elder sister, Sal, died as a result and it is inferred in the story that Josie's illness may also be a serious side effect of the procedure. Despite the dystopian features Ishiguro implies in this society, Josie enjoys a privileged lifestyle in a large family home surrounded by fields and farmland. Josie is often ill and her schooling is all virtual. She only has limited personal interactions with other lifted children at scheduled social activities and also with her closet friend and neighbour Rick, an unlifted child with whom Josie has a loose plan to be together with in adulthood. Living with Josie's family allows

¹²³ Ishiguro, K. *Klara and the Sun* (Faber & Faber, 2021), 58.

Klara to learn more about human relationships and dynamics and she applies this knowledge to her relationship with Josie in her constant striving to be a better AF.

As Josie's health continues to decline Klara develops a sun religion she believes could heal her, appealing to it directly for a cure and offering it the 'sacrifice' of the Cootings Machine Klara believes is responsible for the pollution that blocks out the sun. Josie visits the nearest city to have a portrait completed, a process we eventually find out is intended for the building of a three-dimensional sculpture that Klara could assume in the event of Josie's death. From this scene a key question of the novel is raised; if Klara can imitate Josie in every respect, can she become her? With Josie's subsequent recovery rendering the issue moot, the novel finishes with her away at college and a redundant Klara waiting for the slow 'fade' on her solar battery to complete.

***Klara and the Sun* as Speculative Fiction**

Ishiguro has incorporated many different genres into his novels, from fantasy in *The Buried Giant*, the English aristocratic novel in *The Remains of the Day* and the detective genre in *When We Were Orphans*. Atwood has commented that Ishiguro likes to 'experiment with literary hybrids, and to hijack popular forms for his own ends, and to set his novels against tenuous historical backdrops.'¹²⁴ Many of his works include multiple genres, *Never Let Me Go* for example is more obviously a speculative and dystopian novel, but it also incorporates elements of the Gothic, the satirical and the Bildungsroman categories too. Like *Never Let Me Go*, *Klara and the Sun* falls neatly into the speculative fiction category. As other works in the genre do, it considers the possibility of an imagined world and contemplates what such a landscape would mean and what it may reveal about our present selves. The novel also aligns with Atwood's definition of speculative fiction as the strong-AI presented is certainly considered a scientific possibility.¹²⁵ Other genre elements exist within the novel too, including SF, dystopian and young adult fiction, with events in part four of the book, where Josie sits for her 'portrait' with Mr Capaldi, approaching a Frankenstein-esque Gothic parable.

¹²⁴ Margaret Atwood quoted in: Thomas, *Science Fiction and Speculative Fiction*, 168.

¹²⁵ See page 25 of this dissertation.

Unlike many speculative fiction and SF authors, Ishiguro does not world-build in his novels, there is remarkably little detail given about the settings he creates. In his analysis of Ishiguro's attraction to speculative fiction, Chris Holmes writes that he 'plays a keen game of genre Jenga', looking to see 'how much architecture of novel-making you can remove, how much distance can you put between the character and their floating world and still have a standing, recognizable form.'¹²⁶ With so many of the partial details Klara tell us about her world, the reader is painfully aware of the difference between what she sees and what is actually happening. For example, at the beginning of the novel when sitting in the department store window, Klara observes a homeless man and his dog who were so still 'it was obvious they had died, even though the passers-by didn't know it.'¹²⁷ The next morning with the sun pouring on to the street outside, Klara notices that the man and his dog were in fact not dead, but alive and well due to the sun's 'special nourishment'¹²⁸ bringing them back to life. The reader understands that logically the man and his dog must have been alive all along, but perhaps due to her programming or perhaps because of her robotic nature, Klara is unable to make this connection.

Ishiguro pays more attention to character-building than world-building, but his focus is largely on relationships and the relational. Ishiguro says himself that the reason so many convincing characters in culture often failed to move him was because they:

Didn't connect to any of the other characters in an interesting human relationship...this next thought came regarding my own work: What if I stopped worrying about my characters and worried instead about my relationships?¹²⁹

It is through relationships and interactions with others that Ishiguro reveals what it means to be human in this novel, on this occasion using a robot as the lens to examine the complicated nature of human relationships and the themes this raises.

¹²⁶ Holmes, C. Kazuo Ishiguro's Thinking Novels in Seigneurie, K. (ed.) *A Companion to World Literature* (John Wiley & Sons, 2019), 5.

¹²⁷ Ishiguro, *Klara and the Sun*, 50.

¹²⁸ Ibid, 52.

¹²⁹ Ishiguro, K. (2017). *Nobel Lecture* [Online] Available at: <https://www.nobelprize.org/prizes/literature/2017/ishiguro/lecture/> (Accessed 01 August 2022).

Klara as Narrator

While many authors return to explore the same themes again and again, Ishiguro extends this to his narrators; the mild bamboozlement Klara finds in the world is reflected in Stevens, the butler in *The Remains of the Day*, and his efforts to understand his new employer. Klara's single-mindedness in finding a cure for Josie is echoed in Ryder, the concert pianist narrator of *The Unconsoled*, who believes one perfect performance may be the key to repairing an old family wound. Ishiguro has made a motif out of unreliable narrators such as Stevens and Ryder, as well as Masuji Ono, the ageing painter in *An Artist of the Floating World*. Ishiguro has commented that he wants his narrators 'to be unreliable in the way most normal human beings are. We all represent ourselves in a particular way to others and then we reinterpret and unpack what people say.'¹³⁰ On the one hand Klara can also be considered an unreliable narrator, she is naïve owing to her lack of experience and she is also an outsider; an AF living in a human world. But in one sense she is a completely reliable narrator, very literally observing and cataloguing what she sees and experiences, noticing everything but with varying ability to distinguish between the significant and the insignificant.

At the department store at the beginning of the novel Klara describes her environment, talking about things such as the 'red shelves' and the 'front window' in an almost stoic manner. Her manner of speaking is very literal and precise, for example when she views sees a block of city houses she describes that:

There were six of them in a row, and the front of each had been painted a slightly different colour, to prevent a resident climbing the wrong steps and entering a neighbour's house by mistake.¹³¹

This precision extends to her description of emotions and on one occasion when Klara hears Josie crying she describes the sound with beautiful meticulousness: 'not only was her voice loud, it was as if it had been folded over onto itself, so

¹³⁰ Miller, S. (2021). *Why Nobel Prize Winner Kazuo Ishiguro Thinks his New Book is Like Charlie Daniels' 'Devil Went Down to Georgia'* [Online] Available at: <https://www.oregister.com/2021/03/08/why-nobel-prize-winner-kazuo-ishiguro-thinks-his-new-book-is-like-charlie-daniels-devil-went-down-to-georgia/> (Accessed 01 August 2022).

¹³¹ Ishiguro, *Klara and the Sun*, 241.

that two versions of her voice were being sounded together, pitched fractionally apart.’¹³²

There are many markers that identify Klara as different from humans; she does not eat, faces a wall at night, she cannot smell and sometimes the people she encounters become partitioned into boxes in her vision, an idea Ishiguro describes as cubist.¹³³ She has a very formal and slightly stilted way of talking, asking ‘so Josie would wish me to be present’¹³⁴ rather than ‘you would like me to be there’ or ‘it’s very nice to meet Rick’¹³⁵ rather than ‘it is nice to meet *you*.’ There are countless examples of her speaking in this manner once she moves to Josie’s house, but interestingly in the opening chapters in the department store Klara’s speech with the other AFs is very different. In an exchange with Boy AF Rex, Klara is admonished by him for taking too many of the sun’s rays for herself:

“Because of you,” Boy AF Rex said, “I’m going to become weak by evening.”

“You’re making a joke,” I said to him, “I know you are”¹³⁶

It seems that for Klara, when talking with other AFs the gap between ‘you’ and ‘I’ is bridgeable, but in conversation with humans it is completely impenetrable, either because of her programming or because of any nature, personality, or consciousness she may possess. The chasm between AFs and human beings is also acknowledged in the people Klara encounters, with one minor character telling her: ‘one never knows how to greet a guest like you...after all, are you a guest at all? Or do I treat you like a vacuum cleaner?’¹³⁷

Within Ishiguro’s narratives of tight constraint, he often asks ultimate questions about human experience, sometimes stripped back to ask what being human ever was to begin with. *Klara and the Sun* asks further questions regarding human nature and human distinctiveness; is there anything unique about humans? Could AI ever be considered as equal to human beings or even superior? In the analysis of the novel in the next chapter I examine what this speculative world tells us about

¹³² Ibid, 236.

¹³³ Miller, (2021), *Why Nobel Prize Winner...*

¹³⁴ Ishiguro, *Klara and the Sun*, 88.

¹³⁵ Ibid, 83.

¹³⁶ Ibid, 4.

¹³⁷ Ishiguro, *Klara and the Sun*, 191.

how strong-AI may connect to the idea of the *imago Dei*, and consider what, if anything, it can tell us about our own reality and possible future.

Chapter Four: AI and the *Imago Dei* in *Klara and the Sun*

The examination of the scientific and literary history of AI in the previous chapter identified several of the ethical and theological questions surrounding the subject. *Klara and the Sun* examines some of these issues, particularly those related to human distinctiveness, and this chapter uses the novel as a vehicle to examine the *imago Dei* in the context of AI. This will include considering if Klara qualifies as strong-AI and if so, does she possess consciousness? Other questions that will be contemplated include is Klara religious? Can AI in this novel become indistinguishable from a human being? In whose image are AIs made? These points will be considered where relevant in context of the substantive, functional, relational, and eschatological models of the *imago Dei* discussed in Chapter Two. Before beginning this analysis it is important to reiterate that I am not proposing to read the novel as ‘real’ in anyway, but instead take the ‘imaginative potential’ Cynthia Wallace discusses to examine what it would mean for the *imago Dei* if Ishiguro’s thought experiment were to play out.

Klara as AI

When considering Klara as AI we start from a relatively weak position textually. Ishiguro does not give any detail on how the leap in technology required to create AIs has come about and we do not know the AI strategies employed in creating or programming them. Klara makes no reference to her origins beyond telling us she is a young, or new, AI. Unlike Adam in McEwan’s *Machines Like Me*, who by the novel’s end is seemingly familiar with the entire cannon of human knowledge, Klara does not demonstrate the eidetic intellect often associated with robots. The only indication that she has this functionality at all is when she completes the roboticist Mr Capaldi’s test to determine how well she has really learnt Josie’s nature. In this scene Klara is able to answer questions ‘running to over a hundred digits and symbols’¹³⁸ and by her own admission she tells us ‘having grasped its

¹³⁸ Ishiguro, *Klara and the Sun*, 265.

central purpose, I no longer needed to give it much attention.’¹³⁹ It is likely that we never see this side to Klara during the rest of the novel as it is not necessary in fulfilling her quest to protect Josie. We are told by the Manager that Klara’s real skill lies in her excellent observational abilities and her capacity to ‘absorb and blend everything she sees around her.’¹⁴⁰ This is demonstrated in her understanding of the subtleties and contradictions of human relational dynamics and interaction, such as when she comes to realise that emotionally there can be pain alongside happiness.¹⁴¹ She can reflect on the behaviour, tone, and body language of those she observes and often offers her thoughts on the possible motivations, intentions, and desires behind what she witnesses.

There are examples in the novel of humans treating Klara poorly, for example when a woman mistakenly thinks she is going to watch a play she shouts at the AF ‘first they take the jobs. Then they take the seats at the theatre.’¹⁴² The reader recognises that this comment is unpleasant, but Klara is not bothered at all, her rejection or acceptance by the stranger has no bearing on her primary mission to save Josie. The only sense of self we get is in a brief exchange with Chrissie after the plan for Klara to replace Josie in the event of her death is revealed. While raising her arms in the air to indicate her body, Klara asks: ‘If I were to continue Josie, if I were to inhabit the new Josie, then what would happen to...all this?’¹⁴³

Klara is clearly advanced AI that leaps and bounds ahead of the abilities of robotics and research in present times, but would she qualify as strong-AI? Margaret Boden believes that AGI would exhibit motivation, emotion and cognition¹⁴⁴ and Klara can tick the box in each of these categories. There is ample commentary throughout the novel that Klara is completely motivated by her desire to save Josie, to the extent that she is willing to sacrifice part of her vital components to do so. She claims to experience a range of emotions including

¹³⁹ Ibid.

¹⁴⁰ Ibid, 42.

¹⁴¹ Ibid, 30.

¹⁴² Ibid, 319.

¹⁴³ Ibid, 281.

¹⁴⁴ Boden, *Artificial Intelligence*, 48. Her inclusion of emotion as an essential element of AGI is interesting as initially AI research did not pay much attention to this area, believing intelligence did not require it. It was not until Marvin Minsky published his ‘mind as a whole’ theory that emotion was given real consideration in the field.

happiness¹⁴⁵ sadness¹⁴⁶ and excitement¹⁴⁷ as well as possessing the ability to read the emotions of others. If we take cognition to mean the mental process of acquiring knowledge and understanding through thought, experience, and the senses, we find Klara again has some success. She repeatedly refers to herself as a thinking creature, and the human characters in the novel also reference her as a being with cognitive ability. Even without context Klara can view snapshots of people's lives from the department store window and make observations about human behaviour, with the Manager telling her 'You never miss a thing, do you?'¹⁴⁸

Of these strong-AI criteria Boden believes emotion to be the most challenging for AI to achieve, alongside language and creativity. Within language Klara has strong capabilities with excellent speech recognition and fluent response, although she does occasionally make mistakes, lacking the word for drones and calling them 'machine birds'¹⁴⁹ for example. She understands grammar and can alter her usage depending on whether she is interacting with another AF or a human being; Klara makes sense to those around her. Discussed at length later in this chapter, she demonstrates complex creativity when she develops her own religion in isolation from other characters in the book. This is not the first time Ishiguro has considered the creative element in his artificial beings in a theological context, the clones in *Never Let Me Go* were encouraged to make art, an exercise we find out near the novel's conclusion was to prove they had souls and gain stronger rights and better treatment for them.¹⁵⁰

The Question of Consciousness

Consciousness is often where the distinction between man and machine is made. To be 'conscious' can be construed in different ways; the functional view that includes states such as being awake or asleep or acting deliberately or unthinkingly, and phenomenal consciousness that considers sensations such as pain

¹⁴⁵ For example: Ishiguro, *Klara and the Sun*, 53 and 84.

¹⁴⁶ For example: *Ibid*, 52 and 131.

¹⁴⁷ For example: *Ibid*, 51 and 77.

¹⁴⁸ *Ibid*, 30.

¹⁴⁹ *Ibid*, 81.

¹⁵⁰ Ishiguro, K. *Never Let Me Go*, (Faber & Faber, 2009), 238.

and qualia,¹⁵¹ the introspectively accessible phenomenal facets of our mental lives. The existence of qualia in any form is somewhat of a metaphysical mystery, it is the part of consciousness David Chalmers calls ‘the hard problem’¹⁵² with the philosopher Jerry Fodor stating that ‘nobody has the slightest idea how anything material could be conscious.’¹⁵³

AI can tackle consciousness in two ways; through ‘machine consciousness’ or analysing it in computational terms, without modelling.¹⁵⁴ Klara can be considered to possess the most basic level of consciousness, that of being able to perceive and express self-existence, the ‘I’ to the ‘non-I’. She refers to herself as a conscious entity several times, including in the scrapyard at the end of the novel where she tells us that even though her memories are beginning to overlap, ‘I remain conscious.’¹⁵⁵ Boden states that truly intelligent AGI would have functional consciousness, meaning it would be able to focus on different things at different times, be able to deliberate, self-reflect and generate creative ideas,¹⁵⁶ a brief Klara fulfils. Indeed, it is not too much of a jump to imagine a real future where functional consciousness is realised in AI as it can be understood in terms of information processing.

Key figures addressing AI and phenomenal consciousness include Daniel Dennett, Paul Churchland and Aaron Sloman. Dennett has described the processes of the brain as ‘virtual’ or ‘Joycean’ machines, the former a reference to the virtual worlds found in computer systems. He believes that ‘anyone or anything that has such a virtual machine as its control system is conscious in the fullest sense, and is

¹⁵¹ Qualia can be defined as: ‘A philosophical term for sensory experiences that have distinctive subjective qualities but lack any meaning or external reference to the objects or events that cause them, such as the painfulness of pinpricks or the redness of red roses. The term is virtually synonymous with sense data.’

Oxford Referencing, (2022). *Qualia* [Online] <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100357499> (Accessed 26 Aug 2022).

¹⁵² See Chalmers, D. J. (1995). Facing up to the problem of consciousness, *Journal of Consciousness Studies*, 2: (3), pp.200-19.

¹⁵³ Jerry Fodor quoted in: Deacon, T.W. (2011). *Consciousness is a Matter of Constraint* [Online] Available at: <https://www.newscientist.com/article/mg21228406-300-consciousness-is-a-matter-of-constraint/#:~:text=IN%20A%201992%20issue%20of,anything%20material%20could%20be%20conscious> (Accessed 01 August 2022).

¹⁵⁴ Boden, *Artificial Intelligence*, 109.

¹⁵⁵ Ishiguro, *Klara and the Sun*, 301.

¹⁵⁶ Boden, *Artificial Intelligence*, 109.

conscious because it has such a virtual machine.’¹⁵⁷ While we can consider Klara to be in possession of functional consciousness, we cannot tell if she possesses phenomenal consciousness. This is partly because it is so hard, if not impossible, to define in terms of human experience, but also due to the lack of textual evidence one way or the other. Dennett states that machines could have just as much consciousness as human beings, although he does not believe this to be desirable as they would be ‘precisely as autonomous as we are. And we are very dangerous.’¹⁵⁸

From this analysis I would determine that Klara qualifies as strong-AI who would pass the Turing Test. In addition to fulfilling the conditions laid down by the experiment, Klara also has physical capabilities that seem remarkable for a robotic being. The technology she is built on largely avoids Moravec’s paradox, her movement around Ishiguro’s world is seemingly unimpeded unless on very rough or uneven terrain. While her vision does not work in a way identifiable with the human eye it is highly functional outside of instances when she becomes overstimulated, occasions that cause her field of vision to become segmented into boxes that sometimes cause navigation and perception difficulties. We know that Klara’s model has already been superseded by the B3 AF and she is far from the end point of human robotic advances in this world; she is potentially a point on the journey to AI that is indistinguishable from the human. What, if anything, does this mean for the understanding of human distinctiveness in theological anthropology? Attention now turns to consideration of Klara as AI within each of the substantive, functional, relational and eschatological interpretations of the *imago Dei* discussed in Chapter Two.

The Substantive Model and Klara

In Chapter Two it was noted that the substantive view has generally fallen out of favour among theologians, but to ensure thorough coverage of the topic and acknowledge its historical importance it is certainly worthy of consideration here. The model locates a quality inherent in human beings that highlights their

¹⁵⁷ Dennett, D. *Consciousness Explained*, 2nd Ed. (Penguin, 1993), 281.

¹⁵⁸ McNeil, T. (2020). *Our Brains, Our Selves: Daniel Dennett* [Online] Available at: <https://now.tufts.edu/2020/09/02/our-brains-our-selves> (Accessed 01 August 2022).

distinctiveness from the rest of creation, for many centuries it was believed to be human intellectual ability. Following the Aristotelian philosophical model, this uniqueness separated man from the rest of creation as the only 'rational animal'. The emergence of strong-AI such as Klara has changed this world and humans are faced with an 'other' who has equal, or even superior, intellectual capacities, seemingly exploding the idea of human uniqueness found in rationality. One caveat that should be applied here is that while the results may look the same, the 'thinking' processes behind AIs are almost certainly different to that of human characters, even within our current processing power we know the human brain functions and reasons very differently to a machine.

If we can eliminate intellectual process such as rationality as the essence that makes human beings unique, are there any other qualities we could examine through the lens of Ishiguro's novel? Theologians including Irenaeus and Thomas Aquinas believed the image of God consisted of the endowment of both a rational mind and a free will, the latter usually considered in terms of human beings and their freedom to make individual choices that are not determined by prior cause or divine intervention. When analysing Klara's free will the great difficulty is we do not know the difference between what she chooses to freely do and what she is programmed to do. The only textual hint Ishiguro gives us comes towards the end of the novel, after Josie has been healed, when the Mr Capaldi visits Chrissie to ask a favour from Klara. He tells her:

There's growing widespread concern about AIs...people saying how you've become too clever. They're afraid because they can't follow what's going on inside any more...they accept your decisions, your recommendations, are sound and dependable, almost always correct. But they don't like not knowing how you arrive at them.¹⁵⁹

This brings me back to the earlier caveat concerning the difference in *how* human beings and machines think, a disparity that in Ishiguro's speculative world is at least partially responsible for the anxieties AI is causing. This passage also informs us that humans have lost some level of understanding and control over their

¹⁵⁹ Ishiguro, *Klara and the Sun*, 389.

creations, limiting what Scott Midson refers to as ‘capacity,’¹⁶⁰ the unique human ability to maintain power and dominion over their inventions.

It would seem reasonable to assume parents would only entrust their children to AF companions who had strict built-in and assured safety features, suppressing qualities such as rebellion or insubordination, and as a result their free will, but things are not necessarily as clear-cut as this. After their initial meetings, Klara longs for Josie to return to the store and purchase her. In the intervening months another child shows interest in purchasing her as an AF, but she is blank and unengaging, despite the girl’s obvious enthusiasm. This is noted by the Manager who later admonishes Klara, informing her ‘it is for the customer to choose the AF, never the other way round.’¹⁶¹ Klara apologises to the Manager responding that she thought for that particular child ‘I perhaps wouldn’t be the best choice.’¹⁶² In this situation Klara has a preference, to become Josie’s AF, and she exhibits behaviour to help obtain what she desires. A further example of insubordination occurs when she completes Mr Capaldi’s test, deliberately eavesdropping on the rest of her party even though she understands this infringement may generate anger towards her. The destruction of the Cootings Machine is a criminal offence, something her accomplice Paul, Josie’s father, points out but the AF shows no moral or ethical consideration towards this element of her plan. These instances, while relatively rare for Klara, demonstrate at least a degree of free will and agency; she has some capacity to act with autonomy and make choices to proceed one way or another.

It is also important to consider any external limitations on Klara’s freedom, including any restrictions placed on her free will by those around here. Josie makes it clear on several occasions that she believes Klara has, and rightly so, the right to be autonomous and make choices. In the department store Josie has the following exchange with Klara:

I don’t want you coming against your will. That wouldn’t be fair. I really want you to come, but if you said, Josie, I don’t want to, then I’d say to Mom, okay, we can’t have her, no way. But you want to come, right?¹⁶³

¹⁶⁰ Midson, *Cyborg Theology*, 54. See Chapter Two for further exploration of this idea.

¹⁶¹ Ishiguro, *Klara and the Sun*, 45.

¹⁶² *Ibid*, 44.

¹⁶³ *Ibid*, 23.

Josie has a sensitivity to Klara's preferences that expands and dignifies her field of autonomy; Klara loves to watch the sun set and Josie arranges for her to do so on several occasions, in one instance risking the anger of her mother if caught. While Josie is generally respectful and kind towards her AF, other children could just as easily use contempt and violence to restrict their AFs autonomy. We find out near the end of the novel, when the department store manager finds Klara in the scrapyards, that 'things didn't go as well for Rosa as they did for you,'¹⁶⁴ referencing the AF Klara shared the store window with. Even Chrissie, who begins the novel with little feeling towards Klara, will not allow Mr Capaldi to reverse engineer her to understand how AFs think, telling him 'Klara deserves better. She deserves her slow fade.'¹⁶⁵

Ishiguro, as always, does not provide us with any firm answers concerning robotic free will in this world. On the one hand, all of Klara's behaviour throughout the novel can be viewed as acts of devotion towards Josie, the examples of disobedience and the destruction of the Cootings Machine included. We could interpret these actions as part of her programming, to protect the child she is chosen by at all costs. This view has the useful consequence of maintaining Asimov's 'three rules' but also denies Klara free will. I would argue that in as much as Klara possess functional consciousness, she also has functional free will and autonomy; throughout the novel she appears to make her own choices but there is no metaphysical self, in the Cartesian sense, beneath her synthetic skin.

The Functional Model and Klara

The functional model finds the *imago Dei* in human activity or agency in the created world, gifted upon humanity by God as explained in Genesis 1:26. Klara is first and foremost a product of human endeavour, absolutely defined by her utility. She is a tool for human fulfilment who in this instance is designed to be a companion to a child who may otherwise be lonely, a role she completes with care and dedication. As strong-AI I would argue that Klara does have at least a degree of agency, demonstrating a comprehensive capacity for autonomous thought,

¹⁶⁴ Ibid, 399.

¹⁶⁵ Ibid, 390.

reasoning and action, even within the constraints of her role as an AF. AI and robots in literature have an established tradition of high functionality, from Čapek's forced workers to the androids developed by humans to be used as servants in Phillip K. Dick's *Do Androids Dream of Electric Sheep?* What is unusual about Ishiguro's novel is that the AI is replacing an *emotional* function; Klara is supplying something that would normally be provided by another human being. Ishiguro raises the question that if AI can replace humans in their emotional and relational lives, can they replicate or replace the human in their entirety?

I discussed in Chapter Two concerns the functional interpretation raises around the exploitation of nature and a potential fear around human replacement. Both factors are pervasive throughout the novel; pollution in this world is rife and evidently a much deeper crisis than the one we face today. Klara tells us that in her view from the department store the 'pollution became so bad that, even from the magazines table side, I could no longer see the gap of sky'.¹⁶⁶ Later she describes how after four days of continuous pollution she could feel herself weakening¹⁶⁷ from lack of sunlight. Moravec's prediction that computers will one day displace humans from essential jobs has been realised, with many 'post-employed'¹⁶⁸ people who have found no other work to move on to after the period of the 'substitutions'.¹⁶⁹ We are not told any detail about what this process involved but if we recall the lady at the theatre who challenged Klara, she informed us that AI took 'all the jobs.'¹⁷⁰ The lack of employment opportunities has clearly caused enormous social fractures in society and things appear to be getting tangibly worse. We are told of hundreds of adults and children who are squatting in a building in the city that is due to be cleared with no 'reasonable plan regarding their relocation'.¹⁷¹ By the novel's end we find out that Melania, the housekeeper, has left Josie's family home and is now living in a community because of her concerns about safety.¹⁷²

¹⁶⁶ Ibid, 39.

¹⁶⁷ Ibid, 40.

¹⁶⁸ Ibid, 314.

¹⁶⁹ Ibid, 252.

¹⁷⁰ Ibid, 318.

¹⁷¹ Ibid, 316.

¹⁷² Ibid, 83.

We learn that Josie's father was an engineer before the substitutions and now lives in a community of post-employed people. He mentions others in the group who had 'careers far grander than mine' indicating that many professional jobs are now carried out by AI and robotics. The human intellectual ability, so important in the substantive interpretation of the image, is no longer valued in this world, with Chrissie emphasising this forced redundancy when she asks Paul 'how is it right no one can make use of you?' given his 'unique knowledge' and 'specialist skills.'¹⁷³ The functional abilities of AI to replace human beings in their work has reduced the dignity of many, and while life must be incredibly challenging for many of the post-employed, there is a flip side. Paul claims that the substitutions were:

the best thing that happened to me...made me take a completely fresh look at the world, and I really believe they helped me to distinguish what's important from what isn't.¹⁷⁴

For some in this world, the substitutions will be a form of liberation from the toil of working life to the pursuit of higher aspirations. The clones in *Never Let Me Go* served a similar functional role in unshackling humans from the frailties of the human body, despite the multiple ethical and societal challenges both developments present.

From one perspective Klara falls very much into the functional category; for all her complex abilities she has them to fulfil her role as an AF. Humans here have made life in their own image using their powers as created co-creators, a notion that is very functional in nature. Mr Capaldi claims that Klara can replace Josie in her entirety, not just as a 'copy'¹⁷⁵ or facsimile of the original, but is there textual support for this claim? It is more than possible that Klara could replace Josie in a purely functional way, performing the same tasks and bearing her physical resemblance, but as will be explored later in an examination of the relational and eschatological models, there is more that needs to be considered to allow a full answer to this question.

¹⁷³ Ibid, 252.

¹⁷⁴ Ibid.

¹⁷⁵ Ibid, 210.

The Relational Model and Klara

The relational view finds the *imago Dei* in human-divine and human-human relationships. In this novel Ishiguro concentrates on the existence of relationships rather than focusing on the differences between humans and AFs. Klara clearly has two-way relationships with the other characters in the novel, most notably Josie. For Klara this is a very straightforward dynamic, telling Josie that it is her duty to be her best friend.¹⁷⁶ Josie's relationship with her AF changes over time, she shows many emotions towards Klara including concern, annoyance, and even slight contempt. In response to Klara's statement that she is her best friend Josie corrects her, replying 'you're my AF. That's different. But Rick, well, we're going to spend our lives together.'¹⁷⁷ While Klara may not have a full understanding of the dynamics behind their relationship, Josie is in no doubt that human-human and human-AF relationships are very different. After Josie's health recovers Klara is eventually moved from the bedroom to the utility room, all without complaint or grievance. When Josie leaves for college she thanks Klara for her service and suggests this will be their last meeting: 'I guess you may not be here when I get back. You've been just great, Klara. You really have.'¹⁷⁸ Chrissie is initially slightly cold towards Klara but warms to her as the novel progresses, telling her: 'Josie's become very fond of you. And if I may say so, so have I'¹⁷⁹ Josie's father, Paul, is an 'expert engineer'¹⁸⁰ and initially treats Klara very much as a machine. He also becomes more friendly as the novel progresses, eventually helping her to destroy the Cootings Machine. Rick, Josie's neighbour and boyfriend, is initially suspicious of Klara until he understands her motivations are the same as his; to protect Josie's wellbeing. On several occasions Rick shows great kindness and understanding towards Klara, such as when he challenges a boy at a gathering of lifted children who wants to throw her across the room.¹⁸¹

¹⁷⁶ Ibid, 75.

¹⁷⁷ Ibid, 55.

¹⁷⁸ Ibid, 301.

¹⁷⁹ Ibid, 89.

¹⁸⁰ Ibid, 283.

¹⁸¹ Ibid, 109.

While we can establish that Klara can form and develop relationships, this alone is not enough to fulfil the relational interpretation of the *imago Dei* as discussed in Chapter Two. Barth's relationality for example suggests that personhood and agency are required, a 'Thou' who can respond freely in the Godhead. In Ishiguro's world AI has unquestionably altered humans, changing their relationships, their employment and their very bodies, but discrete boundaries exist between the two species, recognised in Josie's correction to Klara: 'You're my AF. That's different.'¹⁸² An equivalence that allows for human-AI and AI-AI encounters is presented by Ishiguro, but can we also locate a divine-AI relationship here? To answer we first need to address the question of whether Klara can be religious¹⁸³ using textual analysis to examine any practices or belief that she exhibits such as prayer, sacrifice and pilgrimage.

The theologian Robert Geraci argues that for robots to be accepted into society as equals, they must be able to persuade most people that they are conscious beings. An important component of this for Geraci is the ability to form voluntary religious belief and he states that to 'qualify as 'persons'...and thus enjoy equal status in our society, some of them need to be religious - any by choice, not deliberate programming.'¹⁸⁴ *Klara and the Sun* is a story about many things but the question of faith versus rationality is at the heart of the novel, with Klara's sun religion on one side and the roboticist Mr Capaldi's rationalism on the other.

The sun is of course the glaringly unsubtle metaphor that pervades the novel, beginning with the name of its eponymous narrator Klara, from the Latin *Clarus*, meaning clear, bright or famous. We already know that AFs in this novel are solar-powered and while Klara does not require a large amount of sunlight to function, she loves to bathe in it and watch the sun set and rise. Klara's relationship with the sun is already established by the time we meet her, with an almost immediate description of how she liked to lean her face forward 'to take in as much of his

¹⁸² Ibid, 55.

¹⁸³ It is noted that 'religion' and 'religious' can be loaded terms. Academic scholars including J.Z. Smith, Talal Asad and Russell McCutcheon have highlighted the dominance of Western, elite and Christian values found in many definitions.

¹⁸⁴ Geraci, R. (2007). *Religion for the Robots* [Online] Available at: <https://www.bobcornwall.com/2007/06/robots-and-religion-or-can-robot-have.html> (Accessed 01 August 2022).

nourishment as I could'.¹⁸⁵ Her conversations with other AFs in the store suggest her feelings towards the sun are perhaps shared by them, with the boy AF Rex telling Klara that 'the Sun had ways of reaching us wherever we were.'¹⁸⁶ Klara believes the sun possesses divine abilities to heal or even restore life, such as the occasion she witnesses what she believes to be the sun's resurrection of the beggar man and his dog.¹⁸⁷ After meeting Josie, Klara's view of the sun develops into a relationship that requires active participation on her part; her previous thoughts and observations concerning the sun turn into prayers she completes at sunset as this is when the sun is at its closest to the Earth. Later she creates a temple of sorts in a barn she believes to be the place the sun rests at night, a space Klara treats with reverence, considering whether it would be 'discourteous to sit down on this chair, while waiting for the sun.'¹⁸⁸ Once sunlight begins to permeate the barn her prayers turn to pleading and bargaining, asking the sun to 'please make Josie better...[she's] still a child and she's done nothing unkind.'¹⁸⁹

Her destruction of the Cootings Machine is an oblation Klara carries out to draw the sun's attention to Josie's situation in 'some particular and noticeable way.'¹⁹⁰ She makes a second sacrifice in volunteering one of her vital components to destroy the polluting machine, feeling 'great fear'¹⁹¹ at the thought of donating the P-E-G Nine solution. When Paul expresses his willingness 'to grasp at any chance that comes our way,'¹⁹² Klara overcomes this feeling and willingly goes ahead. There is even a hint of theodicy to Klara's beliefs, exhibited when she encounters an angry bull behind a fence while out with Chrissie. She tells us that she feels something more intense than fear, describing a sensation that 'some great error had been made that the creature should be allowed to stand in the Sun's pattern at all, that this bull belonged somewhere deep in the ground far within the mud and darkness.' Her recall of this memory later in the novel evokes her only

¹⁸⁵ Ishiguro, *Klara and the Sun*, 3.

¹⁸⁶ *Ibid.*

¹⁸⁷ *Ibid.*, 52.

¹⁸⁸ *Ibid.*, 217.

¹⁸⁹ *Ibid.*

¹⁹⁰ *Ibid.*, 155.

¹⁹¹ *Ibid.*, 298.

¹⁹² *Ibid.*, 300.

moment of doubt in the sun's abilities, telling us 'I even thought the Sun wasn't kind at all, and this was the true reason for Josie's worsening condition.'¹⁹³

After this brief crisis of faith Klara's hope is restored and eventually the miracle she always expected appears to happen. By this stage Josie is very unwell but when the sun begins to pour through the bedroom window 'illuminating her, and the entire bed, in a ferocious half-disc of orange' she begins to get better. Klara tells us 'the sun's special nourishment proved as effective for Josie as it had for Beggar Man, and after the dark sky morning, she grew not only stronger, but from a child into an adult.'¹⁹⁴ Of course it may be a coincidence that Josie recovers, the important point is that Ishiguro allows Klara and her belief that Josie would heal to win; faith conquers rationality in this story.

This analysis demonstrates that Klara expresses both spiritual beliefs and practices; she has absolute faith that the sun will provide a cure for Josie and develops a set of practices to support this belief including prayer, sacrifice and bargaining. The sun worship Klara exhibits is certainly religious, with similarities to ancient cultures who practiced sun worship including the Nabateans, who built the city of Petra, and the Incas in Peru. If the evolution of human religion is a blueprint, in this speculative world Klara could be the beginning point of an AI religiosity that will also change and grow, possibly at an exponential speed due the AIs advanced information processing skills.

It is somewhat surprising that Klara has developed a sense of religion given that no other character in the novel refers to belief, or otherwise, in any type of God. There are none of the physical remnants of the Christianity we find in *Never Let Me Go*, where the clones use the grounds of an 'old church' to congregate socially and Kathy informs us that the building's caretaker, Mr Keffers, 'is rumoured to be religious' and discredited as outdated by the Hailsham clones. The potential to believe in a deity would perhaps seem an unusual programming choice for an AI built to be the companion of a child, so is there anywhere else Klara's mustard seed of faith could have come from? In *Never Let Me Go* the clones are alert to sightings of their 'possibles', the people from whom they are cloned. On one

¹⁹³ Ibid, 208.

¹⁹⁴ Ibid, 379.

occasion Kathy and Tommy go with their friend Ruth to track a sighting of her possible, staring into her workplace in a manner reminiscent of Frankenstein's monster watching the De Lacey family through the gap in the wall. While they can neither confirm or deny Ruth's origins from their observations, in a burst of feeling she concludes that clones are 'modelled from trash...if you want to look for possibles...then you look in the gutter'¹⁹⁵ and this thought in connection to their beginnings is also reflected in their dreams featuring flying rubbish. Unlike the clones, Klara never expresses any interest or curiosity in how she was created or by whom, so we find no answers here. The helio-centred religion Klara develops implies she is more than a mere machine; she has started to build and populate her own cosmogony, the only worshipper in a world where human beings have lost hope. As a reader, it is difficult to view Klara's acts of faith so differently to those we have perhaps all made.

The Eschatological Model and Klara

The eschatological interpretation shares commonality with the relational interpretation and will now be considered. Pannenberg finds the *imago Dei* in human beings expressed in exocentricity, an openness to the world and to relationships with others that bring transcendence. This interpretation finds direction for human beings at the individual and collective level, propelling them towards realisation of infinite relationship with God in the eschaton. The exocentricity Pannenberg describes looks similar to what Klara expresses throughout the novel; she is driven purely by a motivation outside of herself, namely Josie and her wellbeing. Klara never expresses or pursues any needs of her own and at the end of the novel she has no regrets, reflecting that she has fulfilled her life's purpose; possibly the most unhuman aspect to her. The clones in *Never Let Me Go* share a similar sense of purpose, after making her first donation Ruth tells Kathy that 'It felt right, after all, it's what we're supposed to be doing, isn't it?'¹⁹⁶ A clone who has made all their organ donations is said to have 'completed'¹⁹⁷ and the sense of wholeness the term suggests is reflective of their fulfilment.

¹⁹⁵ Ishiguro, *Never Let Me Go*, 152.

¹⁹⁶ Ibid, 207.

¹⁹⁷ Ibid, 93 & 205.

The clones can extend human life, but Ishiguro takes this one step further with Klara and asks whether a human being can be transferred wholesale into a robotic, raising some interesting points within the eschatological model. When Chrissie asks Mr Capaldi if it is really possible for Klara to continue Josie, he answers in the affirmative, claiming to have proof ‘she’s already well on her way to accessing quite comprehensively all of Josie’s impulses and desires’¹⁹⁸ He states categorically that:

There’s nothing there. Nothing inside Josie that’s beyond the Klaras of this world to continue. The second Josie won’t be a copy. She’ll be the exact same and you’ll have every right to love her just as you love Josie now.¹⁹⁹

Ishiguro is asking the question here: is there something unique and non-transferable inside every human, something that AI research will never be able to understand or replicate? In discussing Chrissie’s fears and doubts around the process of resurrection Mr Capaldi tells her:

We’re both sentimental. We can’t help it. Our generation still carry the old feelings. A part of us refuses to let go. The part that wants to keep believing there’s something unreachable in each of us. Something that’s unique and won’t translate. But there’s nothing like that, we know that now.²⁰⁰

He is the only human character in the novel who unequivocally believes this, alluding to the possible technological replacement of someone he loved as evidence of his point. Josie’s father tells us that deep down he suspects Capaldi may be right and that:

Science has now proved beyond doubt there’s nothing so unique about my daughter, nothing there our modern tools can’t excavate, copy, transfer...Chrissie on the other hand, isn’t like me. She might not know it yet, but she’ll never let herself be persuaded...never mind how well you play your part... never mind how much she wishes it to work, Chrissie just won’t be able to accept it.²⁰¹

It becomes easy to understand why Chrissie would go through with the plan, even with her doubts, when she tells Klara that if she were to lose a second child there ‘would be no other way for me to survive. I came through it with Sal, but I can’t

¹⁹⁸ Ishiguro, *Klara and the Sun*, 210.

¹⁹⁹ Ibid.

²⁰⁰ Ibid, 277.

²⁰¹ Ibid, 224-225.

do it again.’²⁰² At the novel’s end Klara expresses her belief that Mr Capaldi was incorrect and that there is something unreachable about human beings. She tells the manager of the department store that she did not think the process would have worked out very well, not because she ‘wouldn’t have achieved accuracy’²⁰³ but because however hard she tried there would have:

Remained something beyond my reach. The Mother, Rick, Melania Housekeeper, the Father. I’d never have reached what they felt for Josie in their hearts...there was something very special, but it wasn’t inside Josie. It was inside those who loved her.²⁰⁴

This uniqueness is the *imago Dei* we are perhaps all seeking, inside and outside the novel, and Klara, an AI observing humankind locates it in both the relational and eschatological models.

A New Member of the Family?

Margaret Boden asks the question would we - or should we - accept a human-level AGI as a member of our moral community?²⁰⁵ If we answer yes, she concludes this means AI would be the recipient of human moral concern, and their actions as such would be considered morally evaluable. For Herzfeld, the distinction between humans and AI is clear and the latter is an ‘attempt to create a new member of the cognitive family,’²⁰⁶ although she is far from convinced that this is an achievable ambition. I do not believe that the AGI presented by Ishiguro can be considered in such close familial terms; Klara and the other AFs are in no *meaningful* way the same as the human beings in the novel. If we accept the theological Christian tradition that there is a uniqueness within humankind, the doctrine of the *imago Dei* remains regardless of how human-like AGI become.

The computer scientist Christopher Langton suggests that the study of artificial life should be considered part of theoretical biology, which should attempt to define life based on two lineages; those based on carbon and those based on silicon.²⁰⁷ I

²⁰² Ibid, 280.

²⁰³ Ibid, 400.

²⁰⁴ Ibid, 306.

²⁰⁵ Boden, *Artificial Intelligence*, 123.

²⁰⁶ Herzfeld, *A New Member of the Family?*, 235.

²⁰⁷ Langton, C. ‘Artificial Life’ C. Langton, C. (ed.) in *Artificial Life*, (Redwood City: Addison-Wesley, 1989), 1.

would conclude that AFs are best considered as their own species made in the image of humanity, the *imago humanitatis*, but does this mean our *imago Dei* can by extension be shared or reflected in AI like Klara? Herzfeld caveats that ‘we must always be aware of the otherness of any artificial intelligence’²⁰⁸ and warns of the dangers of conflating people and machines. Herzfeld’s ‘othering’ of AI infers to me that she believes only human beings can be in possession of the *imago Dei*. For me, Kathryn Tanner’s proposal that non-human life may share a ‘weak imaging’²⁰⁹ of God seems more likely, allowing a participation with God that perhaps explains the origins of Klara’s innate religiosity.

Ishiguro’s novel is a two-way lens through which we can examine both robotic beings and human beings. Herzfeld writes that ‘our quest for artificial intelligence tells us more about our need to relate, not only to one another, but to the world around us as well.’²¹⁰ Ishiguro himself writes that his ‘book is not an exploration of whether you can upload a human being into a machine, it’s about what that possibility does to our relationship to each other.’²¹¹ Boden also notes the importance of AI in teaching us something about ourselves, believing the most important lesson AI has given us so far is understanding that the human mind is ‘hugely richer, and more subtle, than previously imagined.’²¹²

Having explored Klara as strong-AI and examined how she may intersect with each of the four interpretations of the *imago Dei*, the concluding section of this dissertation will bring these findings together.

²⁰⁸ Ibid, 93.

²⁰⁹ See: Tanner, K. *Christ the Key: Current Issues in Theology*, (Cambridge: Cambridge University Press, 2010).

²¹⁰ Herzfeld, *A New Member of the Family?*, 244.

²¹¹ Miller, 2021, *Why Nobel Prize Winner...*

²¹² Boden, *Artificial Intelligence*, 37.

Conclusion

This section will conclude the dissertation by summarising the key research findings, analysing the limitations of the study, and suggesting opportunities for further research. To reiterate, the aims of this project were two-fold; to consider why speculative fiction is a useful medium for investigating theological questions and concepts and secondly, to consider what understanding of the Christian doctrine of the *imago Dei* in relation to AI could be found in a novel from the genre.

Key Findings

As AGI is currently a hypothetical situation we cannot examine ‘real’ examples of it interacting with each other, with human beings and with wider society. Having employed Ishiguro’s speculative fiction novel *Klara and the Sun* to examine strong-AI and the *imago Dei* I would conclude that the genre allows readers to ask questions about the world around them and investigate a range of different possibilities. The ‘what if’ element to speculative novels permits this investigation, similar to Waggoner’s assertion that the genre allows consideration of unverifiable realities²¹³ and reasserting the notion of literature as a privileged space in which these explorations can be made. It should be noted that the findings below relate only to what is presented in the novel rather than offering any commentary on what the development of AGI in real life would bring.

In the examination of the *imago Dei* in Chapter Two it became evident that the theological expression of human uniqueness, as it appears out of the discourse between AI and robotics research, is not straightforward. Within the substantive model the emergence of AFs in this world has eliminated the historical claim that man is the only ‘rational animal’, the robotic beings presented have superior intellectual capacities to the human beings due to their information processing power. This finding was caveated by the acknowledgment that humans and AFs conduct their thinking in very different ways and the two processes cannot be

²¹³ Waggoner, *The Hills of Faraway*, 9.

viewed as an equivalence. In real world AI and robotics research it is already understood that this is true, with the human brain functioning in a very different way to a machine.

Klara slots into the functional model of the *imago Dei* well, she is the product of human activity in the created world, a tool for human fulfilment. While her skills and abilities are very complex, the reason she has them is to fulfil her role as an AF. Concerns in this interpretation often focus on the door it leaves open to the exploitation of the natural world and the replacement of human beings, both scenarios confirmed in this novel with widespread pollution and legions of 'post-employed' citizens. As well as the anxiety of replacement, Ishiguro also demonstrates through the character of Paul the possibility of human liberation from the constraints of work found in the functional model and AI. Robotic beings in literature have a long history of functionality and while satisfying this interpretation in many respects, I would conclude that Klara's abilities extend beyond the perimeters of what the model can offer.

The relational interpretation is affected by AI in several ways in *Klara and the Sun*. Herzfeld comments that 'as a litmus test of humanity, the image of God not only contributes to defining what it is to be human but also to determining our relationship to the nonhuman.'²¹⁴ AGI such as Klara would add a new dimension to the model, the human-AI, and in the novel there is clearly already a complex dynamic at play between the advantages AFs can bring to human life and the tension in the relationship around replacement and co-existence. Klara demonstrates an ability to engage in AI-AI and AI-human relationships, and with the creation of her own religion she exhibits the nascent possibilities of the type of AI-divine relationship Geraci believes essential to human acceptance of robots as potential equals.

Within the eschatological model of the *imago Dei*, Pannenberg's idea of exocentricity looks very similar to the external motivation Klara expresses throughout the novel, namely her concern for Josie and her wellbeing. When Ishiguro asks if there is something unique and distinct inside each human that could never be understood or replicated by AI, Mr Capaldi's rationalism is

²¹⁴ Herzfeld, *In Our Image*, 15.

contrasted to Klara's faith. Ultimately the AF believes there is something unreachable inside each person, her location of the image is found in the relational and eschatological interpretations, in those that love the individual.

In Klara, human beings have made life in their own image, the *imago humanitatis*, using their power as created co-creators to do so. I would conclude that this does not necessarily absolutely preclude AFs from access to the *imago Dei*, but this could only ever be partial possession, what Tanner refers to as 'weak imaging'.²¹⁵

I aimed for literature and theology to be equal partners in this study and on reflection I believe this equilibrium has been maintained. Becoming familiar with the intersectional research of Noreen Herzfeld was a valuable example of how to achieve balance in intersectional research, any my analysis in Chapter One of the methodologies of Hayles, Nussbaum and Wallace in particular were also very useful in beginning to learn this balancing act.

Limitations

During the early stages of the design of this study my intention had been to include detailed textual analysis of at least two speculative fiction novels, possibly Ian McEwan's *Machines Like Me* or Jeanette Winterson's *Frankissstein* alongside *Klara and the Sun*. All three novels were published within two years of each other, and their shared themes would have added an interesting and revealing 'compare and contrast' element to the study. It quickly became apparent that to keep within the prescribed wordcount this would not be possible with any degree of depth, so for the reasons outlined in Chapter Three Ishiguro's novel became the focus. Despite streamlining the study there are still areas I feel could have been explored further; for example, Klara's agency is touched on in the discussion above, but without spatial constraints I would also have included a thorough investigation of her possession of moral agency and the implications of these findings.

For similar reasons of brevity this study focussed largely on the work of Herzfeld and her writing on the intersection of AI and the *imago Dei*. Featuring the research of another theologian and computer scientist, such as Anne Foerst, would have

²¹⁵ See: Tanner, *Christ the Key*.

added another layer of depth to the study, allowing more exploration of the work of Paul Tillich for example. Other robotics whose voices would have added much to this discussion include Cynthia Breazeal and Rodney Brooks, and expansion could have been made to the sections on Hans Moravec and Alan Turing within the perimeters of a longer study.

The overall analysis was limited in scope by the author's ability to only engage with fiction and secondary sources available in the English language. This placed certain social and cultural constraints on any findings and it would certainly be interesting to investigate this topic from a more global outlook to analyse any differences in perspective.

Areas for Further Research

This section will briefly outline further possible research avenues highlighted by this study. Within genre, the use of a speculative novel worked well for my purposes, but it would be interesting to see further studies that examine AI and the *imago Dei* in other mediums, such as film, television, graphic novels and perhaps even video games, all genres that Eiss believes can also fall within the speculative genre too.²¹⁶ Examination of the image is broken down into discrete sections by many theologians, such as the use of the substantive, functional, relational and eschatological in this study, and these discrete categories perhaps come at the expense of examining the intersection of these viewpoints, an overlap Cortez recognises in his discussion of the 'multifaceted'²¹⁷ view. Additionally, examining AI within the context of the *imago Dei* is but one way of engaging science and theology, AI and robotics could be examined from a wide range of theological angles including cosmogony, free will and soteriology.

²¹⁶ Eiss, *Electric Sheep Slouching Towards Bethlehem*, cvii.

²¹⁷ Cortez, *Theological Anthropology*, 18-21.

Bibliography

- Asimov, I. *I, Robot*, (New York: Doubleday, 1950).
- Atwood, M. *In Other Worlds: SF and the Human Imagination*, (Virago, 2012).
- Atwood, M. *Writing with intent: Essays, reviews, personal prose: 1983-2005*, (New York: Carroll and Graf Publishers, 2005).
- Augustine, *Confessions*, (London: Penguin, 1961).
- Barth, K. *Church Dogmatics: III/2*, trans. J.W. Edwards, O. Bussey and Harold Knight, (Edinburgh: T&T Clark, 1958).
- Boden, M. *Artificial Intelligence: A Very Short Introduction* (Oxford: Oxford University Press, 2018).
- Bray, G. (1991), The Significance of God's Image in Man, *Tyndale Bulletin* 42:2, pp.195-225.
- Brown, D. Davaney, S.G. and Tanner, K. (eds.) *Converging on Culture: Theologians in Dialogue with Cultural Analysis and Criticism*, (Oxford University Press, 2001).
- Carins, D. *The Image of God in Man*, (New York: Philosophical Library, 1953).
- Chalmers, D. J. (1995). Facing up to the problem of consciousness, *Journal of Consciousness Studies*, 2:(3), pp.200-19.
- Clowes, R., Gärtner, K. and Hipólito, I. (eds.) *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts*, (Springer, 2021).
- Cohen, J. *Human Robots in Myth and Science* (London: George Allen & Unwin, 1966).
- Cortez, M. *Theological Anthropology: A Guide for the Perplexed*, (T.&T. Clark, 2009).
- Dainton, B. Slocombe, W. and Tanyi, A. (eds.) *Minding the Future: Artificial Intelligence, Philosophical Visions and Science Fiction* (Springer, 2021).
- Deacon, T.W. (2011). *Consciousness is a Matter of Constraint* [Online] Available at: <https://www.newscientist.com/article/mg21228406-300-consciousness-is-a-matter-of-constraint/#:~:text=IN%20A%201992%20issue%20of,anything%20material%20could%20be%20conscious> (Accessed 01 August 2022).
- Delistraty, C. (2015). Lost Toys and Flying Machines: A Talk with Kazuo Ishiguro in *The New Yorker* [Online] Available at: <https://www.newyorker.com/books/page-turner/lost-toys-and-flying-machines-a-talk-with-kazuo>

[ishiguro#:~:text=Ishiguro's%20views%20on%20religion%20are,things%20that%20we%20make%20up](#)

Dennett, D. *Consciousness Explained*, 2nd Ed. (Penguin, 1993).

De Smedt, Johan. and De Cruz, Helen. (2015). The Epistemic Value of Speculative Fiction in *Midwest Studies in Philosophy*, (39, 1), pp.58-77.

Dorobantu, M. (2019), Recent Advances in Artificial Intelligence (AI) and Some of the Issues in the Theology and AI Dialogue, *ESSSAT News & Reviews*, 29:2.

Eiss, H. (ed). *Electric Sheep Slouching Towards Bethlehem: Speculative Fiction in a Post Modern World*, (Cambridge Scholars Publisher, 2004).

Eliot, T.S. 'Religion and Literature' in *Selected Essays*, 3rd edition (London: Faber and Faber, 1951).

Emerson, A. and Forbes, C. (1984). Living in a World with Thinking Machines: Intelligence Will Not Separate People from Machines, *Christianity Today* 28:2, pp.14-18.

Erickson, M. J. *Christian Theology*, (Baker Books, 1983).

Fergusson, D. (2013). Humans Created According to the Imago Dei: An Alternative Proposal, *Zygon*, 48:2, pp.439-453.

Foerst, A. *God in the Machine: What Robots Teach Us About Humanity and God* (New York: Dutton, 2004).

Geraci, R. (2008). Apocalyptic AI: Religion and the Promise of Artificial Intelligence, *Journal of the American Academy of Religion*, (76:1), pp.138-166.

Geraci, R. (2007). *Religion for the Robots* [Online] Available at: <https://www.bobcornwall.com/2007/06/robots-and-religion-or-can-robot-have.html> (Accessed 01 August 2022).

Görman, U.(ed.) *Creative Creatures: Values and Ethical Issues in Theology, Science and Technology*, (Bloomsbury Publishing, 2005).

Graham, E. *Representations of the Post/human: Monsters, Aliens and Others in Popular Culture*, (Manchester University Press, 2002).

Gregory, A. *Science Fiction Theology : Beauty and the Transformation of the Sublime*, (Baylor University Press, 2015).

Gross, T. (2021). *Kazuo Ishiguro Draws on his Songwriting Past to Write Novels About the Future* [Online] Available at: <https://www.npr.org/transcripts/978138547>

- Haraway, D. 'Cyborg Manifesto', in *Simians, Cyborgs and Women: The Reinvention of Nature*, (Taylor and Francis Group, 1991).
- Hayles, N.K. *My Mother was a Computer: Digital Subjects and Literary Texts*, (University of Chicago, 2005).
- Hayles, N.K. *How we Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, (University of Chicago, 1999).
- Hefner, P. *The Human Factor: Evolution, Culture and Religion*, (Minneapolis: Fortress Press, 1993).
- Heinlein, R.A. and Heinlein, V. *Grumbles from the Grave*, (Del Rey, 1989).
- Herzfeld, N. *Technology and Religion: Remaining Human in a Co-created World*, (Templeton Science & Religion, 2009).
- Herzfeld, N. (2007). A New Member of the Family? The Continuum of Being, Artificial Intelligence, and the Image of God, *Theology and Science*, 5:3, pp.235-247.
- Herzfeld, Noreen. (2005). Terminator or Super Mario: Human/Computer Hybrids, Actual and Virtual, *Dialog* 44:4, pp.347-353.
- Herzfeld, N. (2004). Creating In Our Own Image: Artificial Intelligence and the Image of God, *Zygon*, 37:2, pp.303-316.
- Herzfeld, N. *In Our Image: Artificial Intelligence and the Human Spirit*, (Minneapolis: Fortress Press, 2002).
- IBM. (2012). This is Watson, *IBM Journal of Research and Development*, Issue 3.4: May-June 2012, [Online]
<https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6177717> (Accessed 01 August 2022).
- Ishiguro, K. *Klara and the Sun*, (Faber & Faber, 2021).
- Ishiguro, K. (2017). *Nobel Lecture* [Online] Available at:
<https://www.nobelprize.org/prizes/literature/2017/ishiguro/lecture/> (Accessed 01 August 2022).
- Ishiguro, K. *The Remains of the Day*, (Faber & Faber, 2010).
- Ishiguro, K. *Never Let Me Go*, (Faber & Faber, 2009).
- Jasper, D. *Literature and Theology as a Grammar of Assent*, (Farnham: Ashgate, 2016).
- Jordan, J. M. *Robots* (Cambridge, MA: MIT Press, 2016).

- Jordison, S. (2015). *Kazuo Ishiguro Webchat* [Online] Available at: <https://www.theguardian.com/books/live/2015/jan/16/kazuo-ishiguro-webchat-the-buried-giant-the-unconsoled> (Accessed 01 August 2022).
- Keeley, B. (2015). Speculative Fiction and the Philosophy of Perception in *Midwest Studies in Philosophy*, (39, 1), pp.169- 181, 169.
- Langton, C. (ed.) *Artificial Life*, (Redwood City: Addison-Wesley, 1989).
- Lohr, S. (2022). ‘One Man’s Dream of Fusing A.I. with Common Sense’, *New York Times* [Online] Available at: <https://www.nytimes.com/2022/08/28/business/ai-artificial-intelligence-david-ferrucci.html> (Accessed 28 August 2022).
- McCulloch, W.S. and Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity in *The Bulletin of Mathematical Biophysics*, (5), pp.115-133.
- McFadyen, Alistair. (2012) Imaging God: A Theological Answer to the Anthropological Question, *Zygon*, 47:4, pp.918-933.
- McNeil, T. (2020). *Our Brains, Our Selves: Daniel Dennett* [Online] Available at: <https://now.tufts.edu/2020/09/02/our-brains-our-selves> (Accessed 01 August 2022).
- Mayor, A. *God and Robots: Myths, Machines, and Ancient Dreams of Technology* (Princeton University Press, 2018).
- Mercer, C.R. and Trothen, T.J. *Religion and the Technological Future: An Introduction to Biohacking, Artificial Intelligence* (Palgrave Macmillan, 2021).
- Clareson, T. (ed). *Science Fiction: The Other Side of Realism*, (Bowling Green, OH: Bowling Green University Popular Press, 1971).
- Merril, J. *Science Fiction: The Best of the Best*, (New York: Delacorte Press, 1967).
- Midson, S. *Cyborg Theology: Humans, Technology and God*, (I.B. Tauris, 2017).
- Miller, S. (2021). *Why Nobel Prize Winner Kazuo Ishiguro Thinks his New Book is Like Charlie Daniels’ ‘Devil Went Down to Georgia’* [Online] Available at: <https://www.ocregister.com/2021/03/08/why-nobel-prize-winner-kazuo-ishiguro-thinks-his-new-book-is-like-charlie-daniels-devil-went-down-to-georgia/> (Accessed 01 August 2022).
- Mitchell, M. *Artificial Intelligence: A Guide for Thinking Humans* (Pelican, 2020).
- Moravec, H. P. *Mind Children : the Future of Robot and Human Intelligence*, (Cambridge: Harvard University Press, 1988).
- Muller, V. (ed.) *Fundamental Issues of Artificial Intelligence*, (Berlin: Springer, Berlin, 2014).

- Nussbaum, M. *Upheavals of Thought: The Intelligence of Emotions*, (Cambridge University Press, 2001).
- Nussbaum, M. *Love's Knowledge: Essays on Philosophy and Literature*, (Oxford: Oxford University Press, 1990).
- Oxford Referencing, (2022). *Qualia* [Online]
<https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100357499> (Accessed 26 Aug 2022).
- Oziewicz, M. 'Speculative Fiction' in *Oxford Research Encyclopaedia of Literature*, (Oxford University Press, 2017).
- Pannenberg, W. *Anthropologies in Theological Perspective*, (T&T Clark, 1985).
- Porush, D. *The Soft Machine: Cybernetic Fiction* (London: Routledge, 2018).
- Scholes, R. *Structured Fabulation: An Essay on the Fiction of the Future*, (Norte Dame and London: University of Notre Dame Press, 1975).
- Seigneurie, K. (ed.) *A Companion to World Literature* (John Wiley & Sons, 2019).
- Shelley, M. *Frankenstein; Or, The Modern Prometheus*, (Fine Communications, 2003).
- Tanner, K. *Christ the Key: Current Issues in Theology*, (Cambridge: Cambridge University Press, 2010).
- Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence* (London: Penguin Books, 2018).
- Thewatt-Bates, J. *Cyborg Selves: A Theological Anthropology of the Posthuman*, (Routledge, 2012).
- Thomas, P.L. *Science Fiction and Speculative Fiction: Challenging Genres* (Sense Publishers, 2013).
- Tillich, P. *Systematic Theology, Volume 3*, (Welwyn: James Nisbet & Co, 1964).
- Tillich, P. *Systematic Theology Volume 1*, (Chicago: Chicago University Press, 1951).
- Turing, A. (1950), Computing Machinery and Intelligence, *Mind*, 59:236, pp.433-460.
- van Huyssteen, W. *Alone in the World? Human Uniqueness in Science and Theology*, (Grand Rapids: William B. Eerdmans Publishing Company, 2006).
- Vint, S. & Buran, S. *Technologies of Feminist Speculative Fiction: Gender, Artificial Life and the Politics of Reproduction*, (Palgrave Macmillan, 2022).
- Von Rad, G. *Genesis: A Commentary*, (Presbyterian Publishing Corporation, 1973).

- Waggoner, D. *The Hills of Faraway: A Guide to Fantasy*, (New York: Atheneum, 1978).
- Wallace, C. *Of Women Borne: A Literary Ethics of Suffering*, (New York: Columbia University Press, 2016).
- Walton, H. (ed.) *Literature and Theology: New Interdisciplinary Spaces*, (Farnham: Ashgate, 2011).
- Walton, H. *Literature, Theology and Feminism*, (Manchester: Manchester University Press, 2007).
- Warwick, K. *In the Mind of the Machine: The Breakthrough in Artificial Intelligence* (London: Arrow, 1998).
- Williams, L. (2017). *Margaret Atwood on Christianity, 'The Handmaid's Tale', and what Faithful Activism Looks Like Today* [Online] Available at: <https://sojo.net/articles/margaret-atwood-christianity-handmaid-s-tale-and-what-faithful-activism-looks-today> (Accessed 01 August 2022).
- Wood, G. *Living Dolls: A Magical History of the Quest for Mechanical Life* (London: Faber and Faber, 2002).
- Wooldridge, M. *The Road to Conscious Machines: The Story of AI* (Pelican, 2020).