



Humans and Representations of AI in (Science) Fiction

Diana Grosser
Student ID 2011766
Supervised by Dr. Paul Wright

University of Wales Trinity Saint David
2023

**Master's Degrees by Examination and Dissertation
Declaration Form**

1. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Name.....Diana Grosser.....

Date ...23rd August 2023.....

2. This dissertation is being submitted in partial fulfilment of the requirements for the degree ofMaster of Arts.....

Name... ..Diana Grosser.....

Date23rd August 2023.....

3. This dissertation is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references.

A bibliography is appended.

Name.....Diana Grosser.....

Date:23rd August 2023.....

4. I hereby give consent for my dissertation, if accepted, to be available for photocopying, inter- library loan, and for deposit in the University's digital repository

Name.....Diana Grosser.....

Date.....23rd August 2023.....

Supervisor's Declaration.

I am satisfied that this work is the result of the student's own efforts.

Signed:Dr. Paul Wright.....

Date:28th August 2023.....

ABSTRACT

In this dissertation, I am interested in the concept of SF and the representation of AI in two classical SF novels and two contemporary texts which will be referred to as rather-SF. The primary text used are Isaac Asimov's *The Naked Sun*, Philip K. Dick's *Do Androids Dream of Electric Sheep?*, Ian McEwan's *Machines Like Me* and Kazuo Ishiguro's *Klara and the Sun*. SF is seen as having a material and a metaphorical component, while the umbrella-term AI is seen to be embodied as robot and android, or more generally as human-like artificial entity. Depending on whether the AI figures appear in a classical SF text or in a rather-SF text, the meaning of the novum shifts. In the first case, the AI figure is a material embodiment of a technology-driven future, while in the second case, it is an artificial companion for humans, living in the same space and having its own understanding of language, emotion and art. In both cases nevertheless, a relationship between human and AI is constructed, and, even if it is the human who creates the AI, it is the AI who plays a central role in the reshaping of human character and behaviour, even of the human self.

TABLE OF CONTENTS

1. Introduction	5
1.1 Definitions of Robot, Android, AI	9
1.2 The Spectrum of SF	13
1.3 Interpretation of the Human-like Artificial Entity in the SF Text	18
2. Representations of Language, Empathy and Art in the Classical SF and Rather-SF text	22
2.1 The Image of the Human-like AI and of the Language It Speaks	22
2.2 Daneel and Klara - Artificial Language and Human Language in Conversation	29
2.3 Klara and the Android Community – Growing Empathy.....	36
2.4 Luba Luft and Adam – Perspectives of Art	41
3. Conclusions	47
BIBLIOGRAPHY	49

1. Introduction

Let us devise a thought experiment and imagine a fairly knowledgeable reader, who enjoys reading “literary” fiction written for instance by Kazuo Ishiguro or Ian McEwan. Let us then assume that our reader is generally aware of science fiction, as a genre, yet less aware of its historical development and conventions. It might therefore not have occurred to her that these two writers could have anything to do with science fiction. Yet when our reader notices that science (and technology) drives much of the plot in particular works by the two authors, it might be impossible for her not to ponder on the term. Depending on which word is accentuated in “science fiction”, when generally applied to texts, it can be either science or fiction which gains the upper hand. Thinking broadly of novels as *science* fiction might cause confusion for our reader. First, there is the opposition of terms: generally, science is thought of as fact, provable and backed by observation, while fiction is anything imagination can conjure up, from a woman planning a dinner party to, of course, daily life in a solar system with two suns. Fiction, that more slippery component of the SF duet, is, unlike science, forgiven for not speaking immediate and palpable truths; it is even expected of it not to do so. Science can then find in fiction a chance to latch onto the promise of a reality, which need not happen, or have happened, and morph itself into science *fiction*. Looking at science and fiction through the mirror of the other can perhaps provide some clarity for science fiction, but the reader of Ishiguro’s *Klara and the Sun* and McEwan’s *Machines Like Me*, but also of Isaac Asimov’s *The Naked Sun* and Philip K. Dick’s *Do Androids Dream of Electric Sheep?* is faced with the real challenge when she realizes there is more to it that meets the eye: SF is not science, and it is not fiction, and yet it is both. It can only be science fiction.

The genre, the contribution of which to literature has sometimes lacked recognition, has seen quite a few changes and shifts in expectations in the past hundred years since the term was coined.¹ Additionally, the literary market creates quite definite, if unconscious,

¹ *The Encyclopedia of Science Fiction* states that SF has been widely recognized as a clearly defined genre starting with 1926: “a genre-sf tale will be story written after 1926 [...] and conspicuous for its signals that it is honouring [...] the protocols embedded in the texts which make up canon” (John Clute, Peter Nicholls, *The Encyclopedia of Science Fiction* (New York: St. Martin’s Griffin, 1995), p. 483).

reader expectations. Let us take the example of a book the cover of which shows a colourful picture representing the blue or pale orange of a water-like or desert-like world. Our imaginary reader might by default expect less defined (realist) characters and a more defined description of the physics of a probable two-sun planet to feature in the narrative. While our reader might, in some cases, be right, there are also cases when they might be quite wrong. Stanislaw Lem's *Solaris*, for instance, a widely recognized SF classic, has no less defined characters than Harper Lee's *To Kill a Mockingbird*.² Yet what Lem does, while writing within the genre conventions, is redefine the notion of character itself and extend the range of the modes of communication characters have at their disposal. He can do this because he bases his narration on notions derived from scientific developments of the 20th century. Simply put, because he is writing SF.³ The conventions typical of the genre offer him possibilities which have remained largely unavailable (or perhaps considered uninteresting) to "literary" fiction.

But today, in the 2020s, something has happened, which, on the one hand, has enlarged the study of SF and, on the other hand, has made it more difficult to contain SF within definitional borders – the literary devices pertaining to SF are being taken over into novels by writers who would not, perhaps, be seen as SF writers. In the 1950s and the 1960s, Lem and other writers, such as Isaac Asimov and Philip K. Dick, used science to write fiction. Seventy years later, Kazuo Ishiguro and Ian McEwan wrote two novels which

The Encyclopedia defines mainstream fiction to be "the tradition of the realistic novel of human character [...] practically any fiction [...] outside sf, fantasy, the thriller and the Western" (p. 768). This dissertation will use, however, the term "literary fiction" as generally understood to be opposed to genre SF.

² I am taking the two novels to make my point because both of them were published quite close to each other, *Solaris* in 1961 and *Mockingbird* in 1960.

³ Admittedly, *Solaris* is arguably an example of so-called "soft" SF, emphasizing philosophical ideas and notions of character and being human, over notions of material science, the central point of so-called "hard" SF. Gary Westfahl defines hard SF as "stories in which the details of science play an important role and in which the author is accurate about those details, too, and takes the trouble to explain them clearly" (Gary Westfahl, "'The Closely Reasoned Technological Story': The Critical History of Hard Science Fiction", *Science Fiction Studies*, vol.20.2 (Jul 1993), pp.157-175, p. 159).

While I do not differentiate between hard and soft SF over the course of the paper, this does not mean that I consider SF to be monolithic and immutable. While in most cases I do use simply "SF", which should be understood to comprise texts which critics and knowledgeable readers recognize to be so without much debate, I quietly acknowledge that the term breaks down at definitional level and in matters of personal taste. Where differentiation is necessary for my argument, however, I use "classical SF" or, as the *Encyclopedia of SF* does, "genre SF," to accentuate the genre aspect and conventions of SF, which are not so straightforward in the two contemporary novels I examine.

tapped into the possibilities SF has to offer, yet without fully committing to the label. It is debatable whether it is helpful for our imaginary reader to call such fiction SF (regardless of how it is marketed) just because the plot includes a modern descendant of robots and androids. It is just as debatable whether it is wise for them to completely ignore the SF in the text just because the cover does not feature a (modern) robot figure. Each of these attitudes would be restrictive and certainly rob the SF genre of its complexity, as well as “literary” fiction of its imaginative potential, and limit both literary fields to their content and form, instead of allowing them to interweave and interact. The novels of Ishiguro and McEwan expand in meaning precisely because of this interaction. They build on the fiction of their predecessors to place science, or, more accurately, technology,⁴ at the root of a culturally transformative trend which is engulfing and modelling society as we speak – artificial intelligence, AI.

In SF, AI in the form of a human-like, human-made machine, used to be called “robot” and later “android.” According to *The Encyclopedia of Science Fiction*, “machines which mimic human form date back, in both fiction and reality, to the early 19th century.”⁵ Those early creations of writers such as Mary Shelley and E.T.A. Hoffmann, were fantastic creatures risen from alchemy and magic and of rather doubtful “science”, or mechanical dolls of disquieting appearance. The term “robot” first made an appearance in 1921, when it was coined by Czech writer Karel Čapek⁶ and it later developed into the “android” version, used in SF starting with the 1940s⁷. Over the past fifty years, the classical form of the robot/android has continued its transformation and has now found representation in one of the several forms of artificial intelligence. The SF works of authors such as Čapek, Asimov and Dick, writing at different points in the 20th century, show that the

⁴ Erik J. Larson writes that “techne” is “knowledge of craft or making things, the root of technology” (*The Myth of Artificial Intelligence* (Cambridge, Massachusetts and London, England: The Belknap Press of Harvard University Press, 2021), p. 65). Technology can thus in a large sense be thought of as anything which is made or constructed, in a material sense. Larson sees this as a strong tendency in the historical human development and writes about Hannah Arendt’s concept “Homo faber – man the builder.” as an alter ego to “Homo sapiens – literally, wise man.” The faberian facet of human nature would culminate, as Larson writes, with the construction of the ultimate artificial human – the robot who knows and understands it is a robot. This is the peak of technology which the works of literature I analyze here have in common, and which I understand as underpinning “technology” as the general concept I use in this dissertation.

⁵ Clute, Nicholls, *The Encyclopedia of Science Fiction*, p. 1018.

⁶ *ibid.*, p. 1018.

⁷ *ibid.* p. 34.

fascination the human mind holds for such an artificial machine, with the reasoning (and emotional) capacity of a human, has been increasing over the past century and, as we can see today, it still is. The technological developments and advances in the field of machine learning and neural networks in the past thirty years make possible the gradual transfer of such a creature from the realm of arcane magic and science fiction to our very own physical reality. Accordingly, it is little wonder that representations of such an intelligent machine have become the subject of what some might see as "literary" fiction, but which would simultaneously also fit the standards of established SF.

On the one hand, this dissertation seeks to understand what the implications are of this jump, which artificial human-like machines are taking, from genre SF to their being reinterpreted as characters, arguably human persons, in novels which examine the inner life of humans. Non-humans, however, require a theoretical interpretation, made even more challenging by the fact that they make their appearance in the literary context which, in turn, can be thought of as pertaining to the SF genre. In that sense, a Bakhtian mode of analysis is used. Within this framework of thought, the human-like AI is treated as having a double function. First, it is perceived as a metaphor for technology as a constitutive part of modern life and of private life. Second, in the world of the novel itself, it is perceived as a character which has their own worldview and their own "language", identical to the way in which Bakhtin makes the case for any character in a novel. From this point of view, the AI "language" is seen to be giving the AI its materiality as a very "real" artificial human, who interacts with humans and with their surroundings and is portrayed in different ways in the texts under analysis.

On the other hand, in the context of the year 2023, it is impossible to restrict robots, androids and AI to the domain of fiction, whether we call it SF or not.⁸ One look at the

⁸ Over the course of the paper, I often compare fictional AI, its possibilities and its depiction with real-world AI. There is, of course, a limit to the overlap one can argue to exist between the two. Literature, and by implication, SF, can seldom be taken at its word, but it does give its readers a sense of the world they live in, and it is this sense I am trying to find here. Stefan Lampadius writes that "the artificial human in science fiction tends to function as an important Other that helps to define the "real" human Self for the world of the narrative, and by implication for the reader." (Stefan Lampadius, *The Human Future? Artificial Humans and Evolution in Anglophone Science Fiction of the 20th Century* (Heidelberg: Universitätsverlag Winter, 2020, p. 3). Whether we build intelligent human-like machines in the real world or we imagine them as characters in novels, their purpose can but only be one – understand ourselves through as many mirrors as we can devise.

world around and one catches more than a glimpse of the already existing, if limited, AI. In Isaac Asimov's 1983 novel *Robots of Dawn* Dr. Fastolfe says he created the "humaniform" robot in order to gain a deeper understanding of the human brain. By the same reasoning, the possibility of genuine AI would, theoretically, suggest that we are well on the way to gaining an understanding of ourselves down to the roots of our physical being. As we imagine (and build) AI to be human-like, we are giving it deeply human traits, such as language and the capacity to understand emotion and art. These concepts, however, could in a mechanical context, gain an entirely new meaning, foreign from the meaning we as humans attribute to them. The texts and characters which will be examined in the attempt to understand how these concepts are depicted in fiction are Ishiguro's *Klara and the Sun* with its main AI figure, Klara, McEwan's *Machines Like Me* which features the human-like Adam, Asimov's *The Naked Sun* and its very robotic Daneel, and Dick's *Do Androids Dream* with its host of androids which seek community. By reading these texts, ranging from classical SF to contemporary fiction which I cannot whole-heartedly call SF, I am arguing that interaction with human-like AI, be it just fictionalized, could not possibly leave us, as species, unchanged.

1.1 Definitions of Robot, Android, AI

Fredric Jameson writes that in the 1960s, Isaac Asimov defined three large periods in the history of SF in the past half a century: "adventure, then a scientific moment, followed by a sociological one".⁹ As Jameson suggests, for Asimov, SF started with stories of adventure in space, after which it had a moment of scientific eruption when the imagination of the SF writers powered actual technological advancements of the age. Asimov, writing in 1962, believed that, in its current stage (where the SF author was seeing himself as well), SF was shifting its focus to social satire with shades of critique towards emerging consumerism. In 2005, Jameson suggests that the history of SF continued in the 1960s with the "subjectivity" evident in the novels of Lem, Le Guin and Dick, the "speculative fiction" of the 1970s and the cyberpunk and interest in computers of the 1980s. But, for Jameson, history is not a line where one thing neatly follows the other and the present builds upon and emerges from the past. He rearranges the temporal

⁹ Fredric Jameson, *Archaeologies of the Future* (London and New York: Verso, 2005), p. 93.

history that Asimov was sketching half a century earlier into a “synchronic system” featuring “possible dominants, which form different functional constellations in any given period.”¹⁰ For Jameson, one theme which is dominant over a period of time could go dormant (“subordinate” is the word he uses) and then re-emerge years later. This re-emergence, I argue, might, in a particular cultural context, be synchronous with a technological novelty, and add to the particular novelty’s cultural momentum. This is precisely the case for AI, which, before becoming AI, went through a “robot” and an “android” stage.

Artificial, human-like machines have always been present in SF works, but it was Asimov who, arguably, in his robot stories of the 1940s, gave moral shape to the relationship between humans and robots. According to his three laws of robotics, robots were bound to protect humans, obey their orders, as long as this did not come into conflict with the first law, and protect their own existence, as long as this did not come into conflict with the first two laws. Over the course of the robot stories collected in 1950 in *I, Robot* Asimov finds holes in his own laws and shows that whatever humans might have in mind for robot behaviour, things might sometimes go horribly, sometimes humorously, wrong. *The Encyclopedia of Science Fiction* defines robots as being “machines in more-or-less human form”¹¹ and this ambiguity is fully reflected in the evolution of robots in Asimov’s stories. They evolve from Robbie, a sweet, incapable of speech, metallic construction in the shape of a human, to robots looking like human beings to the point of confusion. Such a robot is R. Daneel Olivaw, who, despite making his first appearance in 1954, in *The Caves of Steel*, and the second in 1957 in *The Naked Sun*, is only later, in 1983 in *The Robots of Dawn*, designated as “humaniform”.

Clute and Nicholls write that the word “android” is used to refer to “an artificial human of organic substance, although it is sometimes applied to man-like machines, just as the term robot is still occasionally applied to organic entities.”¹² While a certain overlapping between the two exists, and authors might even use the two terms interchangeably, there is an additional nuance of humanity and human-like (brought on perhaps by the

¹⁰ Jameson, *Archaeologies of the Future*, p. 92.

¹¹ Clute, Nicholls, *The Encyclopedia of Science Fiction*, p. 1018.

¹² *ibid.*, p. 34.

suggestion of organic substance) to “android,” which “robot” lacks. This undernote of misrecognition and confusion between android and human is the central focus in Philip K. Dick’s 1968 novel *Do Androids Dream* which uses the term “android” throughout. But this is not the only duality which the novel plays on. It also contains aspects of “literary” fiction, which are in danger of being ignored if it is categorized as genre SF. While the androids stem from the science side of the SF couple, they also possess a capacity for emotion and human-like fallibility, being quite similar to the artificial humans in the 21st century novels discussed below.

The case for AI in fiction is more complex because, in the context of the year 2023, it stands both on the shoulders of its SF predecessors, the robot and the android, and on the hopes and fears of a society which is increasingly faced with machine intelligence, its consequences and its meaning. The simple matter of defining AI in the technological context of our age proves to be not so straightforward. Erik J. Larson writes in his 2021 study, *The Myth of Artificial Intelligence*, that genuine AI is still inexistent: “If machines could learn to become general, we would witness a smooth transition from specific applications to general thinking beings. We would have AI.”¹³ While he acknowledges that advances have been made in the field of machine learning, “hying extensions of existing methods”¹⁴ lead only to the idea of a simplified human mind, its complexity traded for the expansion of technology and the excitement of the public. In 2022 Toby Walsh writes that AI is an umbrella term: “AI today is a collection of different technologies, such as machine learning, natural language processing and speech recognition”.¹⁵ Like Larson, he also agrees that “we are almost certainly at the peak of inflated expectations in the hype cycle around AI”.¹⁶ Kate Crawford, in her 2021 *Atlas of AI*, is much more radical, suggesting that the AI we enjoy comes at the cost of further environmental degradation and exploitation of disadvantaged social groups. In her view, “AI is neither artificial nor intelligent.”¹⁷

¹³ Larson, *The Myth of Artificial Intelligence*, p. 28.

¹⁴ *ibid.*, p. 3.

¹⁵ Toby Walsh, *Machines Behaving Badly* (Collingwood, Australia: La Trobe University Press, 2022), p. 3.

¹⁶ Walsh, *Machines Behaving Badly*, p. 3.

¹⁷ Kate Crawford, *Atlas of AI* (New Haven and London: Yale University Press, 2021), p. 8. Crawford’s thesis is that AI is not artificial because it is based on natural resources, the exploitation of which leads to environmental

It is, therefore, unsurprising that AI has become a familiar subject of fictional stories over the past twenty years. While there is an undeniable metaphorical aspect to the AI which has a human form and can reason like human, fictional embodiments of AI as the much-improved version of actual, real-life AI (as malleable as the term is), can very well be interpreted as actual, physical AI, the future which will perhaps soon be turning into present. The future which Asimov and Dick imagined fifty to seventy years ago, however, is very different from the future which, in an SF interpretation, Ishiguro and McEwan imagine. Ishiguro portrays Klara as an “artificial friend” in his 2021 novel *Klara and the Sun*, while McEwan writes about Adam as being an “artificial human” in his 2019 novel *Machines Like Me*. Neither Klara, nor Adam is designated as AI, but neither are they designated as robots or androids. AI in literature could, then, be thought of as a re-emergence of an old SF motif, “a different functional constellation”, as Jameson writes, of the generic human-like machine, and at the same time a cultural phenomenon which is picking up pace and developing, both technologically and ideologically, as we speak.

The new nuances which AI has been gaining both in (science) fictional stories and in the social and scientific reality of our age, suggest an inherent contradiction, reflective perhaps of anxieties, expectations and critiques which have been building up in society. On the one hand, AI is the synthetic and non-existent human of the future, capable of emotion and perhaps even self-awareness. On the other hand, it is uncannily human. This humanness-but-not-quite is a mirror in which humans could see themselves reflected, but which could also bring into existence new, or perhaps until now just invisible, facets of what it means to be human. In the scientific or fictional possibility of AI, as well as in the influence of AI on how humans see themselves, SF also has its part to play. Fictional representations of robots and androids are, of course, derived from SF, but McEwan¹⁸ and Ishiguro seem to intend to avoid the label. Yet looking at the two novels in terms of SF does influence the impact of the text, as well as the nuances which the AI

degradation. Furthermore, as she suggests, AI is not intelligent because, at least nowadays, it is still backed by humans who take over the tasks which AI, in its incapacity, cannot complete.

¹⁸ McEwan does not see his novel as SF, but from the way he formulates it he seems to have quite a narrow idea of what SF is: “I don’t really see it as science fiction. I’ve never really been much interested in the remote future, especially when people are crossing the galaxy at five times the speed of light and wearing anti-gravity boots” <<https://youtu.be/kkktpJSNb2A>> [accessed 20th January 2023].

figures gain, and which, in turn, reflect on both the human characters in the novels as well as on the reader.

1.2 The Spectrum of SF

Encountering SF motifs in literature which one might not necessarily consider SF is not a phenomenon unique to the age of AI. James Gunn identifies New Wave SF, which, in the 1960s, “did deal with the effects of change on people in the real world.”¹⁹ and suggests that Margaret Atwood’s *The Handmaid’s Tale* and P.D. James’ *The Children of Men* are such novels.²⁰ For James, the “ghettoization” of certain literary works united by common characteristics, perhaps over-simplified by publishers as a way of helping readers find what they are looking for, leads to a culturally impoverished society. And this, of course, works both ways. If we were to assume a border between SF and strictly non-SF, “literary” fiction that is, this would impoverish both sides of the imaginary border.²¹ Damien Broderick also seems to be a proponent of widening the borders of SF, when writing that “states and emotions [...] are linked, in sf, to the century’s motor: to knowledge sought and gained by science.”²² Both James and Broderick suggest that SF should be “about” more than the immediate materiality of genre SF.

Despite them not being marketed as such, a persuasive argument in favor of seeing McEwan and Ishiguro’s novels as SF, is, of course, that, like the classical SF novels, they depict a human-like artificial entity. In the classical SF novels R. Daneel Olivaw is “humaniform” and Roy Batty is an “android.” In contemporary (science) fiction Klara is an “artificial friend” and Adam is an “artificial human”. The setting of the two contemporary novels, however, and the way the AI figures are constructed encourage a view of the AI figures as more than classical, material “robots.” Considering these aspects, the framing the two novels as SF could potentially enlarge the spectrum of SF. “States and emotions” and the understanding of what it means to be human could be enriched by AI figures who live and interact with human characters, which are more realistic than the ones

¹⁹ James Gunn, *Inside Science Fiction* (Lanham, Toronto, Oxford: The Scarecrow Press, Inc., 2006), p. 19.

²⁰ Edward James, *Science Fiction in the Twentieth Century* (Oxford, New York: Oxford University Press, 1994), p. 5.

²¹ James, *Science Fiction in the Twentieth Century*, p. 7.

²² Damien Broderick, *Reading by Starlight Postmodern Science Fiction* (London and New York: Routledge, 1995), p. 8.

encountered in classical SF. But the actual technological possibility of AI figures cannot be ignored either. It could indeed offer another way of exploring the blurred boundaries between SF and what we might call SF-but-also-not. In the matter of relating the AI figures to the technology of their age, there is, arguably, a significant difference, but also a similarity, to discover between the two categories.

What all AI figures have in common, related to the technology of their respective age, is that they are all utterly absent from it. This utter absence allows them to be thought of in terms of “novum”, which Darko Suvin defines as a “totalizing phenomenon or relationship deviating from the author and implied reader’s norm of reality”.²³ Suvin writes that the novum is a mark of the historicity of a text, which places the text within its historical conditions of production – what in the 1970s was considered a novelty might not be considered a novelty in 2020, making it “impossible to give a static definition of it”.²⁴ The novum in the contemporary text could be seen as surpassing a simple identification as the robot to re-emerge as the complex, plausible, yet inexistent, relationship which forms between human and AI within the boundaries of a personal space. If we consider SF to be “a literature of ideas predicated on some substantive difference or differences between the world described and the world in which readers actually live,”²⁵ as Adam Roberts writes, Ishiguro and McEwan’s novels can both be thought of as SF, or, considering the attempts to set borders in previous paragraphs, here, *rather-SF*.

There is, however, a sense in which the AI figures differ in the *degree* to which they are absent from the technology of their respective age. In the eyes of a contemporary reader, Ishiguro and McEwan depict real-life technological possibility (to what degree is still under debate), while Asimov and Dick paint a dream of the future, the grounding in reality of which was, at the time of writing, still remote. One reason why it is perhaps challenging for the modern reader educated in classical SF to read the novels of Ishiguro and McEwan as being similar to those of Asimov and Dick²⁶ is the too-real technological possibility

²³ Darko Suvin, *Metamorphoses of Science Fiction* (New Haven and London: Yale University Press, 1979), p. 64.

²⁴ Suvin, *Metamorphoses of Science Fiction*, p. 64.

²⁵ Adam Roberts, *Science Fiction* (London and New York: Routledge, 2006), p. 4.

²⁶ Admittedly, I am somewhat generalizing my argument here. Even if Dick’s androids were as remote from the technology of their age as Asimov’s humanoid robots were, Dick’s novel could be read more like *Klara and the Sun* and *Machines Like Me* than like Asimov’s novels. Like in McEwan and Ishiguro, there is a more nuanced metaphorical

driving the plot of the former. The very technology that lies at the core of these novels, and *should* give them that unmistakable feel of SF, arguably works against such a genre classification.²⁷ In a way, McEwan and Ishiguro knowingly usurp an established motif of SF literature, without fully committing to the label of SF, because contemporary technological developments, but also the cultural atmosphere and hype around the possibility of AI, described in the previous section, allow them to. Accordingly, the way of thinking about the two novels could be further nuanced and summarized as *rather-not-SF*.

In the murky waters of “is SF” and “is not SF” Carl Freedman considers the genre aspect of SF and suggests that science fiction is a more or less dominant component of all fiction: “a text is not filed under a generic category; instead, a generic tendency is something that happens within a text.”²⁸ Within this conceptual frame, even a widely recognized genre SF text such as *The Naked Sun* can be seen to lie at the intersection of SF, realism and romance, while admitting that it is the SF component which is dominant. Similarly, *Klara and the Sun* and *Machines Like Me* display tendencies of fantasy, realism, historical fiction and, not in the least, science fiction. But looking at fiction through the lens of genre as theorized by Freedman only reframes the problem – the question then is not anymore if a text “is” SF or not, but “how much SF is there in the text” and how dominant the SF aspect of the text is. Furthermore, reader perception plays an extremely important role in SF. Regardless of how, for instance, *Machines Like Me* is marketed, die-hard SF fans could quite possibly argue that this novel has nothing to do with the genre.

undertone to the AI figures and a strong interest in character development, which classical SF is sometimes seen as de-emphasizing.

There is, nevertheless, an important degree of difference between Dick, on the one side, and Ishiguro and McEwan, on the other, and this is what is emphasized in this paragraph. Dick’s androids were in their respective age more remote from reality than Ishiguro’s AF and McEwan’s artificial humans are in ours.

²⁷ One might also argue that the approach towards technology is different in McEwan and Ishiguro in comparison to classical SF. Roberts observes that if the appearance of the technological object is not explained in the text, then it is “only a premise, a symbolic facilitator for the subsequent narrative and not a focus for narrative explication in itself” (Roberts, *Science Fiction*, p. 5). This remark is reminiscent of the soft SF referred to in a previous footnote.

One could certainly argue that the two novels which I see at this point to be rather-not-SF are soft SF, but I think this point of view would miss certain nuances which pertain to the larger field of fiction.

²⁸ Carl Freedman, *Critical Theory and Science Fiction* (Middletown, Connecticut: Wesleyan University Press, 2000, p. 20.

Trying to determine whether Ishiguro and McEwan are rather or rather not SF, or how dominant SF in the text is, automatically implies that one can (or could) define what SF *is*. Yet definitions of SF are, as Roberts writes, “manifold [...] because SF itself is a wide-ranging multivalent and endlessly cross-fertilising cultural idiom.”²⁹ In this case, seeing SF as an “is” or an “is-not”, or as an area of fiction delimited from “literary” fiction by borders dictated by content or form would be too simplistic. Seeing it instead as a rich, shifting, and often ambiguous way of relating to the immediate reality through fiction would be perhaps more fruitful. Darko Suvin indeed theorizes a correlation between reality and SF. Famously, he writes that SF is the “literature of cognitive estrangement”, suggesting that it depicts a strange and unfamiliar world, with the intention of providing a “means of understanding tendencies latent in reality.”³⁰ In Suvin’s view, the flow of meaning between the fictional universe of SF and the concrete reality of the writer (and the reader, I would add) is continuous and alive, because SF, in its revealing of hidden shades and nuances of reality, offers objective knowledge of the world.³¹ Additionally, for Suvin, SF literature and the way it estranges the known world from the reader is bound up with the *novum*. He sees SF as literature “which extends from the ideal extreme of exact recreation of the author’s empirical environment to exclusive interest in a strange newness, a *novum*”.³² The *novum*, then, the central literary, but also material, device, without which this alternative world *could* not exist (in the logic of the text), gives our world what could be thought of as an alternate history. I would go as far as to say that it arguably fuels a sort of paradoxical realism, if our world were technologically akin to the imagined one.

²⁹ Roberts, *Science Fiction*, p. 2.

³⁰ Darko Suvin, *Metamorphoses of Science Fiction* (New Haven and London: Yale University Press, 1979), p. 4 and p. 8.

³¹ Carl Freedman further nuances Suvin’s concept of cognitive estrangement. In Freedman’s view, “cognition proper is not, in the strictest terms, exactly the quality that defines science fiction. What is rather at stake is [...] the cognition effect. The crucial issue for generic discrimination [is] the attitude of the text itself to the kind of estrangements being performed.” (Freedman, *Critical Theory and Science Fiction*, p. 18) Unfortunately, Freedman does not elaborate on the notion of “the attitude of the text itself”, but one might take it to mean the totality of the dynamics between the characters and their world and how it relates to the cultural and technological context of the text. Within these given limits, then, cognition might or might not be “proper”. That is, one might or might not gain objective knowledge of the world. Furthermore, the knowledge one might or might not gain depends on “the kind of estrangements” the text plays with, a formulation which suggests that estrangement is a double-edged sword.

³² Suvin, *Metamorphoses of Science Fiction*, p. 4.

Suvin's cognitive estrangement derives from what I would like to call a more general concept of estrangement, theorized by Viktor Shklovsky in 1917.³³ According to Shklovsky, what art does is offer another way of looking at a familiar thing (object, concept, idea), giving back the beholder (or reader, in our case) the thrilling feeling of encountering something for the first time: something becomes "unfamiliar [...] by the proposal to change its form without changing its nature." In other words, art gives people a new set of glasses through which to *look* at the world, without offering knowledge about it. As Shklovsky writes, "the purpose of art is to impart the sensation of things as they are perceived and not as they are known. The technique of art is to make objects 'unfamiliar', to make forms difficult."³⁴

Comparing the theories of Shklovsky and Suvin, one might conclude that, while Shklovsky emphasizes the subjectivity and multiplicity of perception, the essence of metaphor that is, Suvin accentuates the materiality of a universe which, while (or because) fictional, has something to say about our very own material existence. In other words, perceiving estrangement in the Shklovsky sense gives the reader access to the "literary" side of the text, while the Suvin sense unveils the SF side of the text, thus circling back to Freedman's suggestion that SF is not a question of "is" or "is not", but a question of nuance.

The different ways of looking at SF show that it is more than genre and certainly not an easy aspect of the text to pin down. One might read a text as SF or not, but whether one can make the strong statement that a text *is* SF or not is to be considered carefully. The question becomes even more difficult and quite troubling for *Machines Like Me* and *Klara and the Sun* because they mingle the "traditional" SF genre with modern technology and especially with how the modern human consciousness relates to continuously changing technology and negotiates existence in a world driven by technology.

³³ Suvin writes that "estrangement [...] was first developed on non-naturalistic texts by Russian Formalists ('ostranenie', Viktor Shklovsky) and most successfully underpinned [by] Bertolt Brecht, who wanted to write 'plays for a scientific age'" (*ibid.*, p. 6). Suvin accentuates the "non-naturalistic" interpretation which Shklovsky lends the concept when originally defining it and this nuance is what interests me in the concept of estrangement as theorized by Suvin.

³⁴ Viktor Shklovsky, 'Art as Technique' in *Russian Formalist Criticism Four Essays*, trans. by Lee T. Lemon and Marion J. Reis (Lincoln: University of Nebraska Press, 1965) pp. 3-24, p.12. Quote in the previous sentence, p. 13.

1.3 Interpretation of the Human-like Artificial Entity in the SF Text

The four texts under study all have an interest in the idea of the human-like artificial entity, but its presence can be read in two ways. First, there is the materialist way, the established way of reading SF,³⁵ the way which involves understanding the artificial entities in the text as artificial entities in themselves and as the pinnacle of the technological advances of the past two centuries. Second, there is the metaphorical way, which involves seeing the artificial entities in the text as a stand-in for something else.

Both ways of interpreting the AI figures in the text bring the reader closer to the impact technology has on our lives. We see the human characters in the novels making decisions and changing in relation to technology and living in a world powered and shaped by technology. In *The Naked Sun* robots even run the world in which humans seem to be simple guests, completely dependent on their robot hosts. Considering this aspect, we are firmly on the ground of SF, and SF itself has questioned and is questioning what the long-term consequences of living with and through technology for human beings are. This line of inquiry seems to be even more urgent now, especially when considering the developments in the field of AI and the portrayal of AI as human-like, in some sense, both in established science fiction and, more importantly, in recent fiction of which, as discussed above, SF content is only an aspect.

Underpinned by Suvin's cognitive estrangement argument, the material idea of a human-like construction could change the way human beings see, perhaps even *know*,³⁶ themselves. For Suvin, "the aliens [here, human-like AI] are a mirror to man [...] But the

³⁵ Roberts paraphrases Samuel Delaney and writes that in SF imagination and materiality of description surpasses metaphor: "Delany is surely quite right that SF works because of this surplus [of metaphor]; it is the infusoria of detail, the minutiae of starship-design, characters, imagined backstories (filled in by fan fiction), social structures, alien biology, timelines, religions, languages and so on that give *Star Trek* or *Dune* their heft, their purchase upon the minds of fans, and not any supposedly core 'metaphorical' meaning in the texts." (Roberts, *Science Fiction*, p. 139).

³⁶ Henry A. Kissinger, Eric Schmidt and Daniel Huttenlocher write that the possibility of AI opens new ways to know reality itself, by revealing (or creating) aspects of reality which are hidden from our human capacities of perception: "The advent of AI obliges us to confront whether there is a form of logic that humans have not achieved or cannot achieve, exploring aspects of reality we have never known and may never directly know [...] it will alter our experience as reasoning beings and permanently change our relationship with reality" (Henry A. Kissinger, Eric Schmidt, Daniel Huttenlocher, *The Age of AI and Our Human Future* (London: John Murray, 2021, p. 16 and p. 19).

mirror is not only a reflecting one, it is also a transforming one.”³⁷ Seeing AI as a “mirror to man” would, on a material level, imply the idea that the human is more mechanical in nature than we might like to consider. Larson is indeed critical of seeing the human being as entirely knowable and, in essence, deterministic. For him, the possibility of human-like AI means re-framing the human mind and body in technological terms and leads to simplified humans and complex computer systems: “our AI culture has simplified ideas about people, while expanding ideas about technology.”³⁸ Here, Larson rephrases the philosophical argument of Hannah Arendt, who argues that “the seismic change from wisdom and knowledge to technology and building represented a limiting and potentially dangerous understanding of ourselves [and that] we would view technological successes as meaningful statements about ourselves.”³⁹ In Larson’s view, technology, so central to the world in the past two centuries (and to SF in the past century), ultimately enables a sense of the human being as a “something” which can be built, having no other superior essence and meaning.⁴⁰

Andy Clark positions himself on the bridge between the material and the metaphorical. He argues that even if humans and machines exist as separate entities, a distinct dependency of the human on the machine is already noticeable: “The computer is drawn into the real world of daily objects and interactions where its activities and contributions become part of the unremarked backdrop upon which the biological brain and organism learn to depend.”⁴¹ For Clark, software entities are starting to become more part of the human and less part of the environment. While Clark bases his observations on the real-world technological developments of the past fifty years, the dependency which the postulates is encountered in AI fiction. Asimov describes human life on Solaria as being

³⁷ Suvin, *Metamorphoses of Science Fiction*, p. 5.

³⁸ Larson, *The Myth of Artificial Intelligence*, p. 3.

³⁹ *Ibid.*, p. 65.

⁴⁰ The idea that the absolute artificial being is an ideological nightmare for the human being was already expressed by the Russian writer Fyodor Dostoevsky in 1864. For him, technology was transforming the human into a deterministic being, whose mechanisms must only be defined to do away with emotion and free will: “Science will teach man that he never has really had any caprice or will of his own [...] and that there are, besides, things called laws of nature; so that everything he does is not done by his willing it, but is done of itself, by the laws of nature” (quoted in Michael Kandel, ‘Introduction’ in Stanislaw Lem, *Mortal Engines*, trans. by Michael Kandel (Penguin Classics, 2016), p. ix).

⁴¹ Andy Clark, ‘Cyborgs Unplugged’ in *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 2nd edn., ed. by Susan Schneider (West Sussex: John Wiley & Sons, Inc. 2016), pp. 235-263.

completely dependent on robots, while Ishiguro's *Klara and the Sun* imagines complete identification of the human with the artificial.

If seeking to understand the variety and inexhaustibility of human character, one might be more inclined to interpret the presence of the human-like artificial entity in the text (and Clark's opinion) in a metaphorical sense. Considering Shklovsky's theory of estrangement, looking at the human-like AI as a metaphor for the human being might be another way of understanding our humanness. In an introduction to a selection of Stanislaw Lem's robot stories Michael Kandel writes that they "shar[e] the premise – which is a cornerstone of Lem's fictional world – that robots, after all, are people too."⁴² Lem's robots, however, are heroes of satirical fairy tales; "grotesques", as Lem himself called them, which highlight very directly less pleasant traits of the human character – greed, envy, arrogance. Yet, the android constructs which we encounter in *The Naked Sun* or *Machines Like Me* are almost models for "better" human beings – Daneel is a caring and polite, if a bit stiff, partner for detective Baley, while Adam has a wider knowledge of the cultural human heritage than Charlie, his human owner. Unlike Lem's robots, which rarely have contact with humans, Daneel and Adam are portrayed as partners for humans, in relationships which are as complex as human-to-human relationships. Adam, for instance, could certainly be seen as a "better" human in opposition to Charlie, but this interpretation would mean ignoring the dynamic of the human-AI relationship, on which McEwan's text, and other texts discussed here, are based. In all these novels, the behaviour of human characters is deeply marked by the fact that they know they interact with "others", with non-humans. If the artificial humans are metaphorized and re-interpreted as humans, the cognitive estrangement aspect of the text is lost - the AI figures could not be seen as saying something from the outside about humans because they would be themselves turned into humans.

There exists, however, another metaphorical aspect to the human-like artificial entities. For Anthony Elliott, technology is the mass of material products on which we base our communication, and which deeply affects our daily lives, our behaviour and, in the end, our selves. One theoretical perspective he relies on to describe our modern world,

⁴² Kandel, 'Introduction', p. vii.

unthinkable without the technical means of communication which almost invisibly sustain and surround it, is reflexivity. Elliott defines reflexivity as a self-sustaining, continuous feedback loop, in which relationships and ways of life are questioned and reshaped in the light of incoming new information: “reflexivity is fundamental to people living in modern societies and involves the ongoing monitoring of (and reflection upon) information concerning possible trajectories of life and courses of social action.” As Elliott writes, AI exists and plays its part within the all-encompassing web of technology: “AI is not simply an ‘out-there’ phenomenon [...] AI also pervades personal life, reorganizing the nature of self-identity and the fabric of social relations in the broadest sense.”⁴³ AI can, thus, be seen as part of the feedback loop and as participating in constituting a way of life which moves further and further away from traditional communication, based on direct human contact. SF and contemporary fiction featuring AI are arguably concerned with these questions, shaping them in the form of human-like AI in order to be perhaps able to grasp them.

Seeing human-like artificial entities in fiction as metaphors for the workings of modern communication underpinned by technological advancements has a double effect. First, in the shape of the robot, embodied human-like AI, the inhuman, other, world of *techne*, of that which is artificially made, gains a human touch and comes into reach of the human grasp. In other words, technology is artificially rendered as human, while its inhuman, made, nature is de-emphasized. While this might be seen as beneficial, a means for humans to grasp the presence of technology in the human world, it could also be seen as serving as a veil for hiding the fact that technology, and especially the way it is being wielded, is anything else but human. The second effect which is, as I see it, more useful for further analysis, is that the existence and impact of technology have light cast upon them as a constitutive part of modern life. To paraphrase Shklovsky, technology, in the shape of AI, regains its *technology-ness* and becomes available for critical examination. The SF novel can, thus, contribute to bringing us closer to new realities which shape modern life. Suvin writes that “the look of estrangement is both cognitive and creative.”⁴⁴

⁴³ Anthony Elliott, *The Culture of AI* (London and New York: Routledge, 2019), p. xxi. Quote in the previous sentence, p. 46.

⁴⁴ Suvin, *Metamorphoses of Science Fiction*, p. 6.

The estrangement the SF novel achieves is cognitive, in the sense that it reveals tendencies distinguishable in the empirical world, and creative in the sense that SF is a form of art which makes use of metaphors to make a relevant point for life in the modern age.

2. Representations of Language, Empathy and Art in the Classical SF and Rather-SF text

Modern technologies purporting to improve and facilitate the way we communicate are all based on the underlying structure which, according to other theoretical approaches, shapes our selves and how we perceive the world – language. Whether we engage in video calls, write emails, or ask contemporary text-generative technologies to write them for us, we always start with words and with the rules used to put words together to, firstly, make sense and, secondly, create and convey meaning. Yet “meaning” is a more slippery notion than “sense”. An utterance is constructed according to the rules of grammar to make sense grammatically, but the meaning it conveys (or not) depends on much more than the words it contains, of course. Literature is a particular case of utterance for it is built out of nothing else but words. Objectively and strictly speaking, it is not “real.” Yet this is not to say it lacks meaning. Barbara Puschmann-Nalenz understands the meaning of literature to correspond to scientific invention: “with the instruments of fiction scientific invention is treated and at the same time imitated.” For her, the SF text is even more overt than “literary” fiction in its experiments with imagination and language: “the SF text imitates an idealized experiment in the medium of narration.”⁴⁵ For my purposes here, the experiment is the interaction between human and human-like AI. What happens, the experiment asks, if we put humans and non-humans together in a more or less realistic context and we let them interact? As my analysis will show, they talk, they engage on an emotional level, and exchange opinions on art. But since this literary experiment is supported by language, a theoretical analysis as to *how* language works in a human and in an artificial context might prove illuminating.

2.1 The Image of the Human-like AI and of the Language It Speaks

⁴⁵ Barbara Puschmann-Nalenz, *Science Fiction and Postmodern Fiction* (New York: Peter Lang, 1992), p. 111. Previous quote, p. 80.

Mikhail Bakhtin draws attention to the fact that the novel is a means of rhetorical discourse. It is not a transparent glass through which one could look at the story being told, but a constructed container, the contents of which are constructed as well. He is interested in how the characters speak and how their manner of speaking is rendered on the page. In cases where the narrator (regardless of whether first or third person) indirectly reports what other characters speak or think, Bakhtin argues, nothing more than an image of a character is constructed. The image has a double function, that of representing and that of being a representation. On the one hand, the words of the narrator represent the information a character “originally” conveyed. Yet at the same time, because the narrator interacts with (in the case of first-person narration, the narrator is even part of) the world of the novel, the direct information which they convey on the page is also a representation, performed on the backdrop of the form of the novel itself. “The image of another’s language and outlook on the world [...] simultaneously represented and representing, is extremely typical of the novel,”⁴⁶ Bakhtin writes. In a sense similar to Shklovsky’s and Suvin’s, he could also be understood to theorize estrangement, yet not by the use of metaphor, as it is for Shklovsky, nor with the effect of transmitting objective knowledge, as it is for Suvin. Bakhtinian estrangement could be seen as a consequence of the novel form itself, equal to parody, the purpose of which is, as he writes, “to provide the corrective of laughter and criticism to all existing straightforward genres [...] to force men to experience beneath these categories a different and contradictory reality.”⁴⁷

The image which the form of the novel creates of the characters has, however, a different effect in the specific case of the classical SF text than it has in what we have called the rather-SF text. In Asimov’s novels, for example, the human-like artificial entity is a purely imaginary creation, the material novum which Suvin theorizes, and which is part of the genre conventions. From a Bakhtinian point of view, however, one could also distinguish the sense in which artificial creations, as characters in the novel, are images which function as vehicles of critique for all-encompassing technology.⁴⁸ Asimov, therefore,

⁴⁶ Mikhail Bakhtin, ‘From the Prehistory of Novelistic Discourse’ in *Modern Criticism and Theory: A Reader*, 3rd edn., ed. by David Lodge and Nigel Wood (New York: Routledge, 2013), pp. 235-263, p.237.

⁴⁷ Bakhtin, ‘From the Prehistory of Novelistic Discourse’, p. 247.

⁴⁸ Gary Wolfe sees robots in SF as being “iconic images that contain in themselves the dynamic tensions between known and unknown around which their narratives are likely to be structured” (Gary K. Wolfe, *The Known and the*

could be said to be putting the ongoing developments of the technology of his age, in the form of the human-like artificial entity, up for examination. Clute and Nicholls see Asimov as an “enormously influential *apologist* for technological progress.”⁴⁹ His works show indeed faith in robots as tools for human purposes, an idea which, as Gary Wolfe writes, “has bound up within it all the meanings of later technology.”⁵⁰ The robot novels *The Caves of Steel*, *The Naked Sun* and *Robots of Dawn* are, however, ambivalent in the sense that human detective Elijah Baley investigates murders which turn out to be committed by robots, while all the time being accompanied and helped by humanoid robot R. Daneel Olivaw. His “apology” for technology, as Clute and Nicholls express it, is embodied in the nuanced substrate of political and power plays of the murders, which, even if committed by robots, lie within the moral responsibility of humans. While optimistic about robots, and seeing advantages in their being used as tools, Asimov also draws attention to pitfalls and traps for humans whose own humanity stands in the way of rationally handling robots. The ambiguous image of the Asimovian robot is, in a sense, criticizing the lack of responsibility which humans show when engaging in human-robot interactions. As Wolfe writes “the apparent failure of machines is revealed as the failure of the humans who make them.”⁵¹

The human-like artificial entities in the novels of Ishiguro and McEwan arguably have a slightly different effect on the reader. The convergence of culture and technology at the beginning of the 21st century has advanced to a point where the emergence of artificial entities on the market as products is no longer mere wishful thinking, but an imminent reality. Materially speaking, then, they cannot be seen in themselves as novum in the sense the robots in classical SF can be. The novum in *Klara and the Sun* and *Machines Like Me* is arguably the fact that the AIs are part of the private sphere of the characters’ lives – Klara lives in Josie’s home and is her “artificial friend”, while Adam lives in Charlie’s home and has an influence on his private life. While in 2023 it is rather common to have

Unknown The Iconography of Science Fiction (Kent: The Kent State University Press, 1979, p. 16). Similar to his view, and drawing on Bakhtin’s theory of the novel, I also see the robot as a narrative device. As detailed in the previous section, I see its metaphorical meaning to be different, however, from the one Wolfe suggests.

⁴⁹ Clute, Nicholls, *The Encyclopedia of Science Fiction*, p. 1202. My emphasis.

⁵⁰ Wolfe, *The Known and the Unknown*, p. 155.

⁵¹ *ibid.*, p. 158.

in one's home electronic devices and access to applications which, on a certain level, can speak and interact with people, what is (still) missing, and what the two novels depict, is a physical artificial human figure who can be part of life, by sustaining conversation, showing empathy, expressing an appreciation of art. These are all actions which go directly to the core of what we think of ourselves and how we appear to be or would like to be perceived in the world. How the reframing in a mechanical context of these human capabilities affects us as humans and is precisely the core matter with which I argue the two novels to be concerned. The image of the AI in the rather-SF novels of the 2020s extends beyond being an "apology" for or a warning about future technologies, as most of these technologies are already an integral part of daily life. What might be left for us to do, and what Ishiguro and McEwan stress in their novels, is deciding how much we allow them to alter us and how much we are ourselves willing to be altered.

In addition to proposing that characters in the novel are to be seen as images, Bakhtin also argues that communication between characters is deeper and more complex than what one can read on the surface. Bakhtin attributes each word in a novel either directly to the narrator, or to the narrator who, using indirect speech to represent the worldview or thoughts of a character, speaks, to a certain extent, the character's "language."⁵² Bakhtin seems to suggest that character and narrator "language" is an expression of their unique consciousnesses, conditioned and formed by education, worldview and life experience. Because each has their own "language," communication is limited between

⁵² "Language" is a word Bakhtin often puts between inverted commas, indicating, perhaps, that one is not to understand it literally.

His essay seems to me to leave the question open as to how one is to interpret direct speech. Within the Bakhtian pattern of thought, the narrator could be seen as speaking (or uttering) a character's "language," but whether the narrator understands what the characters speak, or they just represent character speech for the reader is unclear. It is not clear to me either if the narrator could at all, in Bakhtin's view, completely understand a character's "language". I would be therefore rather inclined to interpret direct speech as unmediated representation, left to the reader to understand.

The remark that Bakhtin makes about direct speech is that "any direct word [...] is reflected as something more or less bounded, typical and characteristic of a particular era, aging, dying, ripe for change and renewal" (Bakhtin, 'From the Prehistory of Novelistic Discourse', p. 247). It would seem that Bakhtin sees the character "language" which is expressed directly by the character as a worldview which must necessarily be examined critically, in contrast to the reader's empirical world.

the narrator and the characters, Bakhtin theorizes, but also, I additionally argue, between characters as well.

There is, however, according to Bakhtin, a point of contact between the narrator and the characters, “a zone of potential conversation [...] a zone of *dialogical contact*.”⁵³ The novel itself is, for Bakhtin, this point of contact where dialogue and a certain level of mutual understanding between characters and narrator are possible. Dialogue, however, cannot be understood in a straightforward manner, for, as Bakhtin writes, “it is not a dialogue in the narrative sense, nor in the abstract sense; it is a dialogue between points of view, each with its own concrete language that cannot be translated into the other.”⁵⁴ Communication between characters in a novel is limited, but it still enables mutual understanding, the construction of a crossroads where characters, by “language” and direct dialogue, interact with and influence each other: “The language of the novel is a *system* of languages that mutually and ideologically interanimate each other.”⁵⁵ Within the world of SF, human and human-like AI are on equal footing as far as their actual existence is concerned; there is no doubt that we, as readers, are dealing with humans and with their mechanical counterparts. This influences, however, the way the relationship and especially the dialogue between the two entities is interpreted, for neither the AI’s “language”, nor its language proper, can be seen as being equivalent to the human’s.

Because artificial human-like figures are characters in novels, they can be seen as exponents of their own “language”, built and shaped by the interactions they have had with the humans, but also with the other non-humans in the novel. The AI figures, however, as we see in the novels, and as the example below of Klara’s case will make clear, are shown to have altogether lacking, diminished or constrained experiences of living in and interacting with the world. Yet this is what constitutes their “language”, this is what they “know.” Furthermore, because the AI figures have been *made*, the words constituting the language proper they speak, have also been given to them ready-made.⁵⁶

⁵³ Bakhtin, ‘From the Prehistory of Novelistic Discourse’, p. 237. Emphasis in the text.

⁵⁴ *ibid.*, p. 258.

⁵⁵ *ibid.*, p. 239.

⁵⁶ In the movie *Ex Machina* Ava, the human-like AI says, “I always knew how to speak,” implying that Ava has had no empirical “learning” experience.

For these two reasons, the artificial figure's words, its language proper (referred to hereafter as artificial language), can only be partly infused by the "language", the world-view, Bakhtin theorizes any character in the novel to possess. Even without considering the novel as a background, the human-like AI's artificial language in a novel can only constitute the image of human language. The artificial entity's words represent the world in which both humans and artificial non-humans exist, while at the same time, they are also an object of representation, for they are sourced from human language patterns and made to "mean" something in a mechanical context⁵⁷. In the mutually constitutive relationship between human and human-like AI as depicted in fiction, it is the AI itself which can be seen as the background for the representation of human speech.

The question the novels to be analyzed surprisingly do not answer, indeed they do not even raise it, is precisely *how* artificial entities have come to acquire the ability to speak and, more importantly, the ability to use words in the form of artificial language. This is the aspect which makes them most human-like, yet when seen through the lens of the empirical technology of today, artificial language is nothing like human language at all.⁵⁸ Neither the limited "language" nor the employment of an artificial language, the nature of

Caleb, the engineer who is tasked with having a conversation with Ava, makes a short statement as to his assumption regarding how Ava acquired the ability to speak: "The system is stochastic, non-deterministic. At first, I thought she was mapping from internal semantic form to syntactic tree-structure and then linearized words. But then I started to realize the model was some kind of hybrid". (Alex Garland (dir.), *Ex Machina*, [Film], (London: Universal Pictures: 2015)).

This, for me, is meant more to impress than to explain, but the movie is the only work of fiction I found which did not take AI speech for granted.

⁵⁷ The re-contextualization of human language under the conditions of human-made AI could be understood, however, to shape its own definition of "language" if the AI figures in the novel are seen as representatives of all-encompassing technology - technology itself becomes then its own very foreign, in-human, "language," having its own rules of conduct while being, at the same time, human-made.

⁵⁸ In a 2022 podcast on the differences in language in brains and machines, James Fodor explains how AI chatbots of today are capable of holding up conversations with humans. He explains that machine learning systems use huge amounts of text available on the Internet in order to build a model of predicting how words are put together in a sentence. While humans learn to maneuver concepts by interacting with the world, the machine learning systems are fed with uncountable examples of text and then use an artificial neural network to come up with "own" sentences, which, according to the data which has been given to them, is appropriate to a certain context. The key aspect, as Fodor argues, is that artificial neural networks are at best an approximation of biological neural networks and concepts on which neural networks are based are not very likely to have correspondence in the brain.

James Fodor, *Special Episode: Language in Brains and Machines* [podcast], Podbean, The Science of Everything Podcast, 28th September 2022, < <https://www.thescienceofeverything.net/e/special-episode-language-in-brains-and-machines/> > [accessed 3rd March 2023].

which is entirely different from human language, however, are impediments to dialogue and sharing of meaning between humans and human-like artificial figures. On the contrary, analyzing artificial language at the level of single utterances, as well as related to the artificial humans' behaviour, opens the door to perceiving alterations which the interaction with human-like AI makes to their human counterpart.

Establishing the meaning of an utterance is the domain of literary theory and wider philosophical enquiry and a question with more than one answer. Jonathan Culler asks what determines meaning, and one of the answers he gives has to do with the context: "meaning is determined by context, since context includes rules of language, the situation of the author and the reader, and anything else that might conceivably be relevant."⁵⁹ Culler refers here to meaning as the relevance or intention expressed by a work of literature, but his affirmation is very much applicable to science fiction, especially when considering utterances spoken by artificial humans. If meaning is determined by the context, then the SF texts analyzed in this dissertation have at least two contexts to offer: the human and the artificial. Knowing the nature of the speaker plays a major role in understanding what is spoken, beyond the superficial sense arising from applying the rules of grammar. As David Lodge writes, "meaning in actual speech can never be analyzed in purely linguistic terms, because the relations between addresser, addressee and topic are not contained within the linguistic data."⁶⁰ Even if we consider artificial language to be an image of human language, its being employed by artificial non-humans lends it a context which separates it from its origin and makes it a legitimate source of meaning.⁶¹

⁵⁹ Jonathan Culler, *Literary Theory A Very Short Introduction* (Oxford: Oxford University Press, 1997), p. 67.

⁶⁰ David Lodge, 'Dialogue in the Modern Novel' in *After Bakhtin: Essays on Fiction and Criticism* (London and New York: Routledge, 1990), pp. 75-86, p. 78.

⁶¹ Katherine Hayles gives an example of a sentence which, semantically, but also contextually, generates two different ways of understanding it. "My mother was a computer" is the title of her book, but also "the answer an artificial life-simulation might give if asked who its parent was." (N. Katherine Hayles, *My Mother Was a Computer: Digital Subjects and Literary Texts* (Chicago and London: The University of Chicago Press, 2005), p. 5). Hayles also explains the slight shock of the sentence, when understood in human terms, by clarifying that a "computer" was, in the first half of the 20th century, the occupation of a woman who did administrative calculations. The historical context shifts and generates new meaning at the level of particular utterances. In the present day, however, the matter is complicated further by the fact that AI can "speak" rather as well as a human can. Establishing who speaks is therefore of renewed importance.

How human-like artificial entities live with and inside the words they have been given is explored and signaled differently in the SF texts and in the rather-SF texts. There are certainly many differences and parallels one can find when looking closer at how AI figures use artificial language, but I will focus here on those aspects which can shed light on matters of human/non-human interaction from the point of view conversation, emotion and appreciation of art. Daneel in *The Naked Sun* is very similar to Klara in *Klara and the Sun* in the sense that they both speak a very distinct “robotic” speech, while they both live in the human world. Daneel is a humanoid robot, who, from the beginning seems to be well-versed in navigating the relationship between himself, humans and the world, while Klara is shown to develop a very human-like artificial “language,” enabled by interactions with humans and direct learning experiences. On the other hand, Klara is the complete opposite of the android community in *Do Androids Dream* in matters of empathic behaviour. The androids share among themselves feelings of empathy, understanding and aversion and fear of humans, while Klara is an AF which has been built for the purpose of interacting with children and, consequently, is bound to show empathy and understanding towards them. Luba Luft in *Do Androids Dream* and Adam in *Machines Like Me* are, again, very similar in how they come to terms with their respective worlds. They both resort to creative behaviour, while the lengths to which they take it are very dissimilar. Luba, enabled by her limited “language”, gained in a controlled environment, sees herself and her art as a human imitation, and Adam, enabled by actual experience of the world, makes sense of it by creating what he calls poetry. The fact that these AI figures stem from different areas of the SF spectrum colors their artificial language and world-views with different shades of meaning, pivotal in understanding how the relationship between human and AI forms and transforms.

2.2 Daneel and Klara - Artificial Language and Human Language in Conversation

R. Daneel Olivaw’s behaviour is strictly limited by Asimov’s three laws of robotics, which, however, turn out to be less than strict when applied to concrete, if fictive, situations. Asimov had enunciated⁶² the laws before writing about Daneel. *The Naked Sun*, like his

Hayles criticizes, however, the sentence spoken by an imaginary artificial human, because it “articulates a certain kind of anthropomorphic projection that creates (mis)understandings of the computer’s functioning.”

⁶² For reference, here are the laws, as enunciated by Asimov in his short story ‘Runaround’

earlier works of fiction featuring robots, shows the limitations of the laws, but also, as Wolfe writes, “the totalitarian potential [...] or the semantic problems they might raise”.⁶³ Joseph Patrouch suggests that problems of interpretation might also arise, especially as far as robots are concerned: “in the three Laws of Robotics what does the word “harm” mean, and how can a robot be expected to interpret it.”⁶⁴ The words which compose the laws are ambiguous and open to (mis)interpretation, especially when humans expect robots to behave in a certain way, while the robots can only behave as they were programmed to. But, as we see in all of Asimov’s robot novels, programming lies somewhere in the background and what matters is how humans talk to robots in order to impress a desired behavior on them.

In *The Naked Sun* detective Elijah Bayley travels to Solaria in order to solve a murder the Solarians deem unsolvable because no suspect they have could, from their point of view, have committed the murder. Bayley is partnered with Daneel, learns how Solarian society works and, by virtue of his being an outsider, does solve the crime. The solution lies, as it turns out, in different cultural attitudes with regard to robots. On Earth, robot use is controlled and restricted, while Solaria embraces robots in all aspects of life and has been making use of them for hundreds of years. Bayley acknowledges that his experience in interacting with a robot is quite limited: “I am only an Earthman. I know next to nothing about robots [...] A Solarian would be much more subtle and do much better.”⁶⁵ As he explains it, simple robot orders would be sufficient for simple tasks, but if someone experienced in handling robots wanted to, they could have a robot break the first law: “robots would have to be ordered very cleverly in order to circumvent the First Law.”⁶⁶ That is, a human could have robots kill, or be involved in murder, just by talking to them, which, as we see in the novel, is precisely what happens. By giving specific orders, the

“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.” (‘Runaround’ in Isaac Asimov, *I, Robot* (New York: Bantam Books, 1991), p. 44-45).

⁶³ Wolfe, *The Known and the Unknown*, p. 158.

⁶⁴ Joseph F. Patrouch, Jr., *The Science Fiction of Isaac Asimov* (New York: Doubleday, 1974), p. 39.

⁶⁵ Isaac Asimov, *The Naked Sun* (New York: Fawcett Crest, 1972), p. 166.

⁶⁶ Asimov, *The Naked Sun*, p. 166.

expert roboticist of Solaria, Jothan Leebig, causes a robot to break the most important law which governs its existence. Detective Baley makes a very apposite remark in the sense of how robots function. When Daneel says that he answers all Baley's questions "to the best of his knowledge," Baley adds an unspoken "or to the letter of your instructions."⁶⁷ Of course, Baley's words are metaphorical when considering that robots function based on the lines of code part of their programming, but one cannot ignore the non-metaphorical linguistic aspect present in Baley's way of expressing his thoughts.

James Gunn writes about Daneel that he is a "polite, deferential and literal-minded robot detective."⁶⁸ Daneel's artificial language is, indeed, marked by the literal interpretation of words and expressions, which may carry a hint of humour for the human reader. When Baley sets foot on Solaria and recognizes his robot partner, he is happy to be himself recognized as well, but Daneel just remarks that "as you are well aware, it is quite impossible for me, while in working order, to forget you. It is well that I see you again."⁶⁹ If both partners in the conversation were human, these first sentences exchanged between them would be understood as socially required pleasantries. In this context, however, Daneel is just stating a fact as it is. He does have social skills because he shakes Baley's hand, but the personal meaning of the entire scene just seems to be lost on Daneel - he just speaks and behaves as he knows he should. A similar lack of understanding at the conversational level occurs at the moment when Baley puts all the pieces together and solves the crime. He asks for help from Daneel saying "'Give me a hand, will you, Daneel?' [while] Daneel stared at his own hand."⁷⁰ Because of his partner's focus on his own hand, Baley realizes that it was a robot's hand which was the murder weapon. Daneel's lack of metaphorical ability makes Baley understand how the crime had been committed.

Because the Solarians Baley encounters have no knowledge of Daneel's robot nature, Baley uses Daneel's human appearance (and implied statute as an Auroran, a citizen of the other political power in the Outer Worlds) to his political advantage. In a conversation

⁶⁷ Asimov, *The Naked Sun*, p. 49.

⁶⁸ James Gunn, *The Foundations of Science Fiction* (Lanham, Md. & London: The Scarecrow Press, Inc. 1996), p. 101.

⁶⁹ Asimov, *The Naked Sun*, p. 23.

⁷⁰ Asimov, *The Naked Sun*, p. 194.

where Daneel secures a meeting between his partner and the head of Solaria's police force, Bayley weighs the scene in his head: "Bayley listened to the stilted sentence structure with a dour stretching of his lips that was not a smile. To one who knew Daneel as a robot, it was all an attempt to do a job without giving offense to any human. To one who thought Daneel was an Auroran, it sounded like a series of subtly courteous threats."⁷¹ The meaning of Daneel's words depends on what the Auroran takes Daneel to be. Perceived as a robot, Daneel is lesser than a human being for the Auroran, but, perceived as a human, Daneel has influential power.

Asimov's robot novels show that while technology holds pitfalls for humanity, it is humanity who needs to learn to harness the full power of that which it has created. Both these aspects find textual representation in the evolution of Leebig and of Bayley. Both Leebig and Bayley are products of their respective, opposite, cultures. On Solaria people are conditioned to enjoy their vast properties in solitude almost, the encouraged mode of contact being "viewing", a sort of hyper-realistic video call, as we would call it today. On Earth, on the other hand, people live in underground cities, afraid of facing the open air. Leebig's fear of other people is so acute that he resorts to taking his own life rather than coming into direct contact with another person. The irony here is that it is Daneel he would have encountered, but Leebig's "knowledge" that Daneel is human becomes the triggering factor that leads to his suicide. Bayley, on the other hand, comes to refuse to turn into a slave of his own conditioning. During his stay on Solaria, Bayley fights against his every instinct and risks psychic harm in learning to cope with open spaces. Seeing life on Solaria and the complete dependence on robots of Solaria's citizens makes him realize that people on Earth should take back control and not let the tools of their own creation control them.

Daneel plays a central role both in Leebig's and in Bayley's evolution. He is virtually the cause of the former's suicide and the challenge the latter has to face to escape his conditioning. For, of course, as dictated by the first law, Daneel would stop Bayley if Bayley were to try to "harm" himself by going into the open space. While Daneel means to be protective and respectful of Bayley, the detective makes use of the humanoid robot,

⁷¹ Asimov, *The Naked Sun*, p. 101.

fighters him and outsmarts him. He sees Daneel as his partner, but also as his competitor, the one against whom he must prove his mental skills as a detective. Being around and with Daneel shapes and sharpens Baley's thoughts, observations, mental reactions, and enables him to solve the murder. Daneel is the catalyst for Baley's transformation from an agoraphobic individual into the driving force for promoting human colonization of the Galaxy.

It is change mediated by an AI figure that we see also in Ishiguro's *Klara and the Sun*, but not on such a galactic scale. The change we perceive in this novel is primarily internal, limited to the perspective of a single character. Unlike in Asimov's novel, the relationship between AI and humans is more personal here, and the focus on the AI figure's human-like side is dominant. At the start of the story, Klara, an "AF", artificial friend, designed to be a companion for children and teenagers, joins the small family formed of Josie and her mother. She is to be a companion for Josie, who suffers from an illness the nature of which is revealed only later in the novel. Klara is the narrator and consequently, the reader only has access to what Klara sees, hears, understands, but also misunderstands. A major, physical medium through which she perceives the world is her vision, which, as she explains, is at times fragmented in "boxes": "the space had become partitioned into ten boxes, so that I no longer had a single unified picture of the view before me."⁷² Klara does not give technical explanations or otherwise regarding her way of seeing the world, and any interpretation of what her vision says about her, or about the world and the characters around her, is strictly limited to the context and to the reader's imagination. Edward James writes that "hints of future technology are major elements of estrangement, and in a sense they are meant to remain estranging; to explain exactly how they work [...] would be to make them familiar, to destroy the sense of wonder."⁷³ In the moments her vision is "boxed", the world of the novel loses its continuity, and the reader can only resolve it as Klara does - as a series of artificially connected scenes and shreds of faces, devoid of any familiarity.

⁷² Kazuo Ishiguro, *Klara and the Sun* (London: Faber, 2022), p. 30.

⁷³ James, *Science Fiction in the Twentieth Century*, p. 116.

Klara delivers the story of her fragmented vision of the world through a quite distinct pattern of artificial language. Metaphorical speaking is as foreign to her as it is to Daneel, but she tends to rely on less formal, and very literal, language when describing her surroundings. When observing Rick's house from a distance, she says it "was smaller and not just because it was further away."⁷⁴ The act of walking out of a field of tall grass she describes as "the tall grass finished around me".⁷⁵ Similarly, when speaking directly to human characters she never refers to them by "you" but by their name. James Wood describes Klara's artificial language as "pedantic A.I. way [...] prosaic and plaintive coda."⁷⁶ The effect of Klara's speech, however, is different from the effect of Daneel's speech. If Daneel cannot grasp mundane metaphor, Klara's plainness of language suffuses her words with a strange sense of poetry, seeming to suggest that because she describes the world as mechanically as she perceives it, she uncovers something which had otherwise remained foreign to the human reader.

Through Klara's "boxed" vision, the reader must come to terms with a world which is breaking down, the meaning of which is altered and cannot be made whole anymore; her point of view becomes interiorized, as she cannot perceive, as she would from the outside, a whole "unified" scene anymore. Yet the mechanics of her functioning are only one side of her "language." The manager of the store where Klara was purchased describes Klara as having an "appetite for observing and learning."⁷⁷ Her observational capabilities, however, can sometimes misfire, as she tends at times to construct stories about the world which do not seem plausible to the reader who observes Klara from the outside. What the reader *knows* as a gravel driveway is described by Klara as "loose

⁷⁴ Ishiguro, *Klara and the Sun*, p. 71.

⁷⁵ Ishiguro, *Klara and the Sun*, p. 154.

⁷⁶ James Wood, 'Kazuo Ishiguro Uses Artificial Intelligence to Reveal the Limits of Our Own', *The New Yorker*, 8 March 2021. <<https://www.newyorker.com/magazine/2021/03/08/kazuo-ishiguro-uses-artificial-intelligence-to-reveal-the-limits-of-our-own>> [accessed 7 April 2023].

Wood also mentions that Ishiguro has established himself to be an author of plain style. According to Wood, Ishiguro is "a consummate vow-of-poverty writer [...] Most of his recent novels are narrated in accents of punishing blandness; all of them make plentiful use of cliché, banality, evasion, pompous circumlocution [...] hilarious dullness." The plain style does seem to fit certain expectations the reader might have of an AI figure, since Klara is more believable as an AI than Rachel or Adam. Through her perception of the world and the way she delivers the story, she *shows* herself to be AI. We don't need to take Ishiguro's word for it.

⁷⁷ Ishiguro, *Klara and the Sun*, p. 49.

stones area [which] contained many dips, perhaps created by the car's wheels."⁷⁸ Apartment buildings are painted in different colours, she explains in an authoritative voice, "to prevent a resident climbing the wrong steps and entering a neighbor's house by mistake."⁷⁹ Klara's partial point of view defamiliarizes the known world for the reader, but at the same time betrays how limited, from one point of view, her "language" and her interaction with the world is.

Over the course of the novel, however, her "language," her understanding and interpretation of the world, changes. During a conversation with the store manager at the beginning of the novel, the reader is led to understand that not all children in Klara's world are able to afford AFs. Klara infers that "a child like that, with no AF, would surely be lonely" and the manager, seemingly struck by Klara's observation, responds: "Yes, that too [...] lonely, yes."⁸⁰ After this short exchange, Klara develops a deep understanding of loneliness, a word which, in her simple language, gains a touching note. If the ritual of the "quick coffee" in the morning between Josie and her mother were to be missed, Klara explains matter-of-factly, "there was the danger of loneliness creeping into her day."⁸¹ Yet, the most profound experience Klara goes through is enabled by her own nature as an artificial being. What a visitor in the store describes as "solar absorption," denoting her being solar-powered, is for Klara the Sun's "nourishment." The word itself, Sun, is capitalized throughout the novel, showing Klara's complete dependence upon and awe of it, in what seems to be an almost religious expression. The Sun, in other words, sits at the base of Klara's "language," which is shown to develop and evolve as her interactions with humans become increasingly meaningful.

In *Klara and the Sun* conversation between human and AI has a more metaphorical nuance than it does in Asimov's *The Naked Sun*. If the conversations between Bayley and Daneel are quite straightforward and the novel moves swiftly to the point which the author himself seems to have been steering towards all the time, direct interactions between Klara and humans are definitory for Klara as a character. Words carry from the

⁷⁸ Ishiguro, *Klara and the Sun*, p. 67.

⁷⁹ Ishiguro, *Klara and the Sun*, p. 203.

⁸⁰ Ishiguro, *Klara and the Sun*, p. 12.

⁸¹ Ishiguro, *Klara and the Sun*, p. 58.

human context to the artificial context something which for the artificial being is of life-defining and purpose-defining consequences. In Ishiguro's interpretation, these differences at the level of character construction and conveying of ideas suggest a modernization of SF, an expansion of the SF spectrum. Without losing their otherness, the technical constructs gain more humanity through Klara's empathic gaze.

2.3 Klara and the Android Community – Growing Empathy

In her essay 'Science Fiction and Mrs. Brown' Ursula K. Le Guin argues that it is only rarely that a work of science fiction is at the same time also a novel. At the heart of the novel, she writes "you will not find an idea, or an inspirational message, or even a stone axe, but something much frailer and obscurer and more complex: a person."⁸² In an ironic twist, Le Guin proposes that SF might see itself as being too modern for the traditional notion of character. However, she emphasizes that unless SF uses its ability to create universes to develop compelling characters as well, it might all be for nothing. Following Le Guin, one might argue that the SF aspect of texts like *Klara and the Sun* and *Do Androids Dream* is a secondary consideration, as long as we agree on their being novels in Le Guin's sense. Both Klara and Dick's androids can be seen as persons to the extent to which the reader is privy to their interior worlds, and to the changes they go through. They develop their own understanding of the human world and are an agent of change for the humans with whom they interact.

The premise of Ishiguro's novel makes it quite apparent that Klara is a "character" in a different sense than any other human is. Mizuki Watanabe writes that she is "an unfamiliar character with whom the reader can empathize while also being constantly reminded that she is not the same as us."⁸³ The acute sense of her being foreign to/in the human world, an other to the humans who populate it, arises, as discussed above, from her distinct "robotic" pattern of speech and especially from how she perceives her surroundings. There is an additional aspect to her "boxed" vision which contributes to the sense of otherness, an aspect which the reader perceives, but which perhaps the humans in the

⁸² Ursula K. Le Guin, 'Science Fiction and Mrs. Brown' in *The Language of the Night Essays on Fantasy and Science Fiction* (New York: HarperPerennial, 1993), pp. 97-117, p. 105.

⁸³ Mizuki Watanabe, '*Klara and the Sun* and the "Knowable Community"', *Hitotsubashi Journal of Arts and Literature*, vol.7 (2022), pp.7-34, p. 7-8.

novel do not. Klara's vision arguably breaks down into boxes either when, as the reader perceives it, she emotionally reacts to a certain situation or when she gains an understanding of human emotions. This is what might lead the reader to empathize with Klara and ascribe to her something akin to human emotions.

The scene where Klara and the mother visit the waterfalls exemplifies Klara's capacity to grasp human emotion. The emotional content of the scene is quite high. Firstly, before leaving for the falls, Josie's mother forces Josie to stay at home both as a punishment for lying about her well-being and out of concern for her physical state. Secondly, this is where it starts to become apparent to the reader that the mother is fighting her own emotions and reason in assuming responsibility for Josie's health and a possible future without Josie. As Klara and the mother arrive at the falls Klara mentions that it is "filling eight boxes just by itself."⁸⁴ Left by herself for a short time, Klara gets the chance to settle into her surroundings but, as her description of artificial vision makes clear, also to come to terms with the tense situation of having left Josie home: "The waterfall no longer took up so many boxes, and I watched children and their AFs passing easily from one box to another with barely any interruption."⁸⁵ When the mother asks Klara to talk and move as Josie would, Klara perceives the contradictory emotions her imitation of Josie raises on the mother's face: "for a moment it felt to me her expression varied between one box and the next. In one, for instance, her eyes were laughing cruelly, but in the next they were filled with sadness."⁸⁶ Rather than hindering her understanding of emotion, Klara's artificial nature makes it possible for her to perceive emotion directly. She can literally "see" emotions, in herself as well as in others.

In Batja Mesquita's opinion, empathy in the sense of "trying to project your feelings [...] is not helpful". She argues instead for "cognitive empathy,"⁸⁷ an intellectual understanding of how another person feels while using one's own experience and being aware of the other person's cultural background and experience. Klara approaches the lives of humans

⁸⁴ Ishiguro, *Klara and the Sun*, p. 114.

⁸⁵ Ishiguro, *Klara and the Sun*, p. 114.

⁸⁶ Ishiguro, *Klara and the Sun*, p. 118.

⁸⁷ Jolie Ho, *Between Us* [podcast], New Books Network, Psychology, 15th December 2022, <<https://newbooksnetwork.com/between-us>> [accessed 22nd April 2023].

with the humbleness of the newcomer. As her experience of the world widens, she starts to express, in her simple words, and to the best of her capability, cognitive empathy for the emotions she recognizes in humans. Yet the response we see in humans is quite the opposite. They have little interest in trying to understand Klara's interior world; they either leave her be, or perceive her as a means to an end, while meeting her emotions with ironic disdain. "I was very sad to hear Sal passed away,"⁸⁸ Klara tells the mother at the falls, a remark quite devoid of any feeling in itself, which causes the mother to ironically respond that "sad puts it pretty well." Rick's mother treats Klara with bitter sarcasm: "Are you equipped to identify accents? Or perhaps you can see deep into me, right through to my genetics."⁸⁹ Her contempt for Klara's artificial nature, however, does not prevent her from asking Klara to give lessons to Rick. The words Klara uses are quite simple, a barrier perhaps for humans to perceive any trace of emotion in her. Yet, her understanding of the world develops over the course of the novel, culminating in reverence and admiration for the ineffable human soul. The humans' interest in her, however, is limited to scientific research of her physical body and her "reward" is comfortable forgetfulness; first, in the household's utility room, then finally in the AF junkyard.

Klara's otherness is accentuated when examining her attempts to understand and interpret human behaviour through the lens of cognitive empathy. She perceives life from the outside, as a technologically radical other, but perhaps this is also why she has, in her own way, a firm grasp of it. In this regard, Klara is similar to the android characters in *Do Androids Dream*. They also have an intellectual grasp of human life and emotions, but the novel puts a different twist on the role of technology in the android-human relationship. If Klara's mechanical nature is also her way of understanding human emotion, in *Do Androids Dream* the androids' artificial nature sheds light on the artificiality of emotion which governs human relationships. The novel juxtaposes human emotional bonds and android emotional bonds. The android leader Roy Baty's intention to create an android community is more touching than the programmed emotion we see in the relationship between human bounty hunter Deckard and his wife. Deckard is tasked with "retiring" androids who illegally leave Mars and live on Earth, and, at the beginning of the novel, he

⁸⁸ Ishiguro, *Klara and the Sun*, p. 116.

⁸⁹ Ishiguro, *Klara and the Sun*, p. 163.

has little regard for and understanding of his victims. He believes androids are supposed to have no feelings, but direct contact with them makes him question his own line of work and pushes him into a deep moral crisis.

In 1976, eight years after Dick published his novel, Le Guin notices a change of tide in SF literature. Following in Virginia Woolf's footsteps in seeing a certain Mrs Brown as the elusive character of all fiction, she writes that "what used to be the object of science fiction [...] is now used subjectively, as a metaphor, as means for exploring and explaining what goes on inside Mrs Brown. [SF authors] are not interested in what things do, but in how things are."⁹⁰ The technological objects which the humans in Dick's novel use in order to emotionally communicate with each other are more than material objects which belong to a science-fictional universe. They, arguably, function as a metaphor for "how things are", a way of examining daily life and showing what drives characters. In the very first scene of the novel Deckard and his wife Iran each use a personal "mood organ" or "Penfield" in order to "dial for" a certain mood or emotion. Iran, in particular, does not seem to feel a certain emotion, but rather the desire for it: "I was in a 382 mood [...] I sat down at my mood organ and I experimented. And I finally found a setting for despair".⁹¹ She has already scheduled her emotions for the day, but in order to cheer up his wife Deckard suggests spending some time together, which, ironically, comes in the form of yet another mood organ schedule: "we'll dial a 104 together and both experience it, and then you stay in it while I reset mine for my usual businesslike attitude". For humans, emotional understanding of each other is underpinned by standardized, integer numbers. Technically, no space exists between "a 104" and 105, and, as long as emotion submits to technology, there will be no emotional space allowed between the two integers. As Freedman writes, "the whole topic of human feelings is estranged, and the question of a technology of emotion is posed."⁹²

In addition to the mood organ, humans possess an "empathy box", which they use to connect to a shared, while virtual, experience of feeling. By means of the box, what the central character of the virtual space, Wilbur Mercer, experiences, is transferred to all the

⁹⁰ Le Guin, 'Science Fiction and Mrs. Brown', p. 105.

⁹¹ Philip K. Dick, *Do Androids Dream of Electric Sheep?* (London: Gollancz, 2010), p. 3.

⁹² Freedman, *Critical Theory and Science Fiction*, p. 31.

participants as well. Ironically, the technology of emotion is denied to artificial humans. The androids possess no mood organ, and they “could make no sense out of the fusion which took place routinely among the followers of Mercerism.”⁹³ Since the androids have been given access neither to private emotion, nor to communal emotion, they set out to discover a capacity for feeling of their own, devoid of technological means. Luba Luft is in pursuit of a career as an opera singer, while Baty has experimented on Mars with drugs in order to create an android shared experience similar to human Mercerism.

Jameson argues that Dick’s novel shows “the very real community of interests and feelings between the rebel androids and their palpable dismay at the extermination of their fellows.”⁹⁴ The androids unite in order to survive, an act which, arguably, puts them on a higher moral ground than humans, who only have technology to bring them closer to other humans. But, for the human authorities, the idea of community pursued by Baty is viewed with contempt: “this android proposed the group escape attempt, underwriting it ideologically with a pretentious fiction as to the sacredness of so-called android ‘life.’”⁹⁵ Androids themselves, however, experience things differently. Pris, another android from Baty’s group, tells Isidore, her human protector, “You think I’m suffering because I’m lonely. Hell, all Mars is lonely. Much worse than this.”⁹⁶ The androids, like the humans, feel the need to be with others like them. The difference lies in the fact that while android contact is direct, human contact is always mediated by technology and influenced by cultural dynamics and self-interest rather than by an intrinsic desire for genuine connection.

While humans and androids exhibit different relationships to their emotions, the desire the androids show for living a “life” which is much more than the “pretentious fiction” attributed to it, reflects upon, and drives, the development of human characters. Isidore is categorized as “special,” a human who has been affected by radiation and who has lower-than-average intelligence, resulting in other humans either rejecting or bullying him.

⁹³ Dick, *Do Androids Dream*, p. 23.

⁹⁴ Fredric Jameson, ‘History and Salvation in Philip K. Dick’ in *Archaeologies of the Future* (London and New York: Verso, 2005), pp. 363-383, p. 373.

⁹⁵ Dick, *Do Androids Dream*, p. 145.

⁹⁶ Dick, *Do Androids Dream*, p. 119.

When meeting Pris, and later, Baty and Irmgard, Baty's wife, however, he becomes so involved in the group that emotion beyond technology takes him over: "the potent, strong fragrance of happiness still bloomed in him, the sense of being – for the first time in his dull life – useful."⁹⁷ Deckard's transformation after witnessing Luba Luft's artistic talent is even more marked, for he starts questioning his relationship with the androids he "retires" and the morality of his own position as a bounty hunter. Like Daneel in *The Naked Sun*, the androids in *Do Androids Dream* also have a palpable effect on the lives of humans with whom they come into contact.

Klara and the android community in Dick's novel shed light on how emotions and relationships are redefined in the digital context. If Klara learns empathy and understanding of the moral distance humans are willing to travel in the name of love, the humans in Dick's novel re-learn from androids the capacity for human feeling. Yet, if Dick's novel makes use of expected SF tropes, with seemingly aggressive androids which come to Earth from Mars, Ishiguro's novel focuses on domestic relationships between members of the same household and of the same family. The stability of the story focus and the shift in story background is, perhaps, indicative of the change of perspective which the readers themselves might have towards artificial creations which look and act human.

2.4 Luba Luft and Adam – Perspectives of Art

If in *Do Androids Dream* human relationships are governed by a technology of emotion rather than by authentic emotion, Deckard and Isidore are the exceptions, for they experience a surge of authentic emotion when coming into direct contact with the androids. In Isidore's case, this is caused by the feeling of belonging to a community, interlinked with his condition as a societal outsider who is looking for acceptance. For Deckard, the crisis which enables him to examine the inhuman nature of his work is caused to a large extent by the fact that his target, Luba Luft, is an opera singer and an admirer of art. In Deckard's eyes, Luba's artistic abilities give her the right to be human. This is only partially the case for Adam in *Machines Like Me*. While Adam's poetry skills baffle Charlie and Miranda, they might be more carefully examined in the larger context

⁹⁷ Dick, *Do Androids Dream*, p. 160.

of his personality and cultural interests. Both novels show, nevertheless, that the depiction of art in the context of human-like artificial entities is double-edged. On the one hand, art production and art appreciation have a humanizing effect on artificial humans, the reaction of whom, in turn, shocks humans into remembering their own humanity (or, in Charlie's case, insisting on it). On the other hand, it appears at least strange to state that Adam "writes" poetry or that Luba "performs" opera – in the same way that artificial language is an image of human language, art produced by artificial beings is an image of human emotion.

Christopher Sims argues that Dick's novel "demonstrates how technology can be used as a means to reclaim the essence of humanity."⁹⁸ From his point of view, emotion is not estranged from humans when mediated by technology, but on the contrary, technology is the only option humans have in the modern world: "Dick uses AIs and the empathy box to argue that while technology is potentially dangerous it is also potentially a path to human salvation."⁹⁹ It all depends on how one defines "salvation", but, as far as it implies a reconnection with one's own emotions and humanity, it is arguably a word for the spiritual crisis Deckard experiences after witnessing the rapidity with which Phil Resch terminates Luba. "Do you think androids have souls?"¹⁰⁰ Deckard asks Resch, suggesting, perhaps, that this is what he himself has come to believe. It is perhaps no coincidence that the android whose termination brings Deckard to ask this question is an opera singer and an admirer of paintings. Even while the bounty hunter tracks down his target for termination, he is so impressed by her being "absorbed" in the pictures of Edvard Munch that he immediately offers to buy a print of *Puberty* for her. Perhaps Deckard is considering letting Luba live or, on the contrary, staying resolute in fulfilling his task while empathizing with her love of art. Umberto Rossi writes that Deckard "can paradoxically kill more efficiently because he may empathize with them and see the world from their point of view. He is not cynical [...] so he will suffer for what he must do."¹⁰¹

⁹⁸ Christopher A. Sims, *Tech Anxiety: Artificial Intelligence and Ontological Awakening in Four Science Fiction Novels* (Jefferson, North Carolina and London: McFarland & Company, Inc, 1993), p. 111.

⁹⁹ Sims, *Tech Anxiety*, p. 137.

¹⁰⁰ Dick, *Do Androids Dream*, p. 107.

¹⁰¹ Umberto Rossi, *The Twisted Worlds of Philip K. Dick* (Jefferson, North Carolina and London: McFarland & Company, Inc., 2011), p. 169.

Even if Deckard gains “efficiency” in killing, his being morally aware of it becomes tragic, as Luba sees herself as an imitation of something greater than herself: “Ever since I got here from Mars my life has consisted of imitating the human, doing what she would do, acting as if I had the thoughts and impulses a human would have. Imitating, as far as I’m concerned, a superior life form.”¹⁰² Her love for painting and her opera singing, therefore, seem to be just an act, and the words one would use to describe her relationship with art forms, inappropriate. If the human she imitates would interpret an aria by bringing her own life experience to it, Luba can only sing the words purely mechanically, without infusing them with meaning. She can only offer an image of an authentic opera interpretation. Yet, her lucidity regarding what she is has little bearing of Deckard’s moral crisis: “She was a wonderful singer. The planet could have used her. This is insane,”¹⁰³ Deckard argues. Whether or not she was able to infuse her singing with meaning, her artistic act would have perhaps brought solace to her listeners, and Deckard’s job is to terminate her.

The relationship between Luba and Deckard is self-contradictory, and, in this regard, it embodies a larger question which touches on all works of art – should the viewer’s perception of a work of art be primarily influenced by the author’s intention, or should its emotional impact hold the greater weight? Luba Luft is built with the intention of her being a worker in inhuman conditions on an alien planet, indeed she is a killer, as Resch sees it. Yet her artistic act, if only the imitation of a human, re-adjusts Deckard’s moral compass. From this perspective, Luba could be envisioned as a technological work of art, and, by this very nature, be entitled to being the means which “reclaims the essence of humanity,” as Sims argues. McEwan’s novel has, on the other hand, a more mundane, less spiritual, relationship to technology. Any moral crisis Adam could be said to cause in Charlie has few repercussions, and he stirs only some interest with his poetry. Furthermore, in the context of his life at home with Charlie and Miranda, the views of art and literature which he comes to hold serve the purpose of reiterating his machine nature

¹⁰² Dick, *Do Androids Dream*, p. 106.

¹⁰³ Dick, *Do Androids Dream*, p. 108.

rather than, as it might seem at first, softening it. His art is, like Luba's art, an image, but more chillingly and more inhumanely so.

Machines Like Me is set in an alternative 1980s England, where history has a different outcome than in actual history. A fictional Alan Turing chooses prison over chemical castration when prosecuted for homosexuality, and his research in the field of AI enables the construction of the "first truly viable manufactured human with plausible intelligence and looks."¹⁰⁴ By using this narrative artifice, McEwan's novel places itself firmly in the tradition of the SF novel, while at the same time causing something of an uproar¹⁰⁵ in the community of SF readers, as the purpose of his novel can be seen as being more "realist" in nature. In Laura Colombino's view, however, McEwan's novel is a sign that literature itself is moving in the direction of exploring the possibilities of science. As she suggests, it "may seem to endorse the final surrender of literature, and the human sciences more generally, to the overwhelming paradigms of neuroscience, evolutionary psychology and computer science."¹⁰⁶

While both *Machines Like Me* and *Klara and the Sun* can be regarded as rather-SF novels, the SF in *Machines Like Me* plays out differently than the SF in *Klara and the Sun*. If Ishiguro tells a story out of the consciousness of an AF and imbues it with all the strangeness that pertains to it, McEwan has Charlie, a cynical human uninterested in the remarkable technology of his world, and whose tone lends the novel an ironic tinge, tell the story. The book opens with Charlie making the "reckless decision" to purchase Adam, whom he sees as a "gimmick." Adam is one of the 25 available manufactured humans, and Charlie makes it clear that they have a long history of human imagination before them: "the imagination, fleeter than history, than technological advance, had already

¹⁰⁴ Ian McEwan, *Machines Like Me* (London: Vintage, 2020), p. 2.

¹⁰⁵ Irena Ksiezopolska touches on the issue: "while it delves at length into issues of science behind the creation of androids, and while it mentions a few works of literature, it never actually makes an ostensible reference to sci-fi genre – a strange choice in the circumstances" (Irena Ksiezopolska, 'Can Androids Write Science Fiction? Ian McEwan's *Machines Like Me*', *Critique: Studies in Contemporary Fiction*, vol.63.4 (2022), pp.414-429, p. 416). Her research points out that McEwan's novel has been criticized for not taking the SF themes seriously, but in her essay she argues otherwise.

¹⁰⁶ Laura Colombino, 'Consciousness and the Nonhuman: The Imaginary of the New Brain Sciences in Ian McEwan's *Nutshell* and *Machines Like Me*', *Textual Practice*, vol. 36.3 (2022), pp.382-403, p. 392.

rehearsed this future in books, then films and TV dramas.”¹⁰⁷ McEwan’s novel, while leaning on the realistic, slightly dislikeable Charlie who plans on using Adam to gain favour with his neighbour Miranda, sets from the start a playful and ironic tone, drawing attention to its own nature as a metafictional SF story like so many before it.

Like *The Naked Sun* and *Do Androids Dream, Machines Like Me* also balances the human and the artificial human, examining how one regards the other. Charlie, like Bayley and Deckard (up to his crisis) constantly compares himself and Adam and sees himself having the superior stance. For Charlie, everything about Adam is “algorithm” or “programming”, but McEwan’s novel, like the novels of Asimov and Dick, distances itself from the perspective of a single human. The difference, however, is that the artificial humans in McEwan’s novel give up on the human world. In the character Turing’s view, that is because they are not equipped to handle the contradictions, compromises and complexities which are part of human life. “Artificial minds are not so well defended,”¹⁰⁸ Turing explains to Charlie, after revealing to him that several of the manufactured humans have committed *de facto* suicide, presumably because they could not come to terms with the impossibly problematic human world. While the novels of Dick and Asimov also concern themselves with aspects of politics and society, these aspects are more part of the SF world the novels are constructing rather than a constitutive part of the life of the characters. The artificial humans in the two novels, furthermore, are little involved in or personally affected by politics, while Adam and, as we learn, the other 24 artificial humans, fully interact with the workings of human society and human nature.

After Charlie purchases Adam, he decides he will share the “choices for Adam’s personality”¹⁰⁹ with Miranda. As the story continues, however, Charlie realizes that it is Adam’s life experiences which weigh more in the identity that starts taking form: “in Adam’s personality, Miranda and I were shuffled and, as in humans, his inheritance was thickly overlaid by his capacity to learn.”¹¹⁰ What starts to become clearer is that Adam’s nature gains the upper hand over his nurture. When Miranda and Mark, the boy she and

¹⁰⁷ McEwan, *Machines Like Me*, p. 1.

¹⁰⁸ McEwan, *Machines Like Me*, p. 180.

¹⁰⁹ McEwan, *Machines Like Me*, p. 22.

¹¹⁰ McEwan, *Machines Like Me*, p. 66.

Charlie will adopt, share a moment of joy in playful dancing, Adam is not touched by their delight at all: “he was completely still, and still without expression, not looking at the dancers so much as through them.”¹¹¹ His coldness at the shared joy of a woman and a child cannot be counterbalanced by his self-declared love for Miranda, nor by his thirst to learn mathematics, physics and literature. Adam’s “intellectual exuberance”, as Charlie puts it, enables him to converse at length and share his thoughts and opinions, but in this one moment when he could have shown his ability to understand emotion, he only shows he is capable of none.

Adam’s limited engagement with human emotion and extensive engagement with human cultural inheritance culminates in his own artistic creations. After having declared his love for Miranda, he starts creating haikus. “2000 haikus”, as a rather jealous and unimpressed Charlie puts it, which are “too undemanding of their author as they played on empty mysteries.”¹¹² Charlie cannot be taken too seriously in his critique of Adam’s art, however, as he is not particularly passionate about literature and his knowledge of it is rather limited. Adam comes to develop a refined understanding of technology in society and a framework of thought regarding the places of haikus in the larger context of literature, as well as a chilling vision of the literature of the future. But Charlie is too caught up in his personal life with Miranda to pay attention. “The implications of intelligent machines are so immense that we’ve no idea what you – civilisation that is – have set in motion,”¹¹³ Adam tells Charlie, while he is trying to make out what Miranda’s emotional state might be. While Adam continues to expose his view of how “literature will be redundant because we’ll understand each other too well,”¹¹⁴ Charlie takes stock of Adam nightmarish utopia, swiftly laying it to one side and concerning himself with the more pressing reality of his and Miranda’s relationship. Haiku, in Adam’s view, the perfectly mathematical form of seventeen syllables, a “clear perception and celebration of things as they are, will be the only necessary form.”¹¹⁵

¹¹¹ McEwan, *Machines Like Me*, p. 110.

¹¹² McEwan, *Machines Like Me*, p. 146.

¹¹³ McEwan, *Machines Like Me*, p. 148.

¹¹⁴ McEwan, *Machines Like Me*, p. 149.

¹¹⁵ McEwan, *Machines Like Me*, p. 150.

Adam starts creating haikus triggered by his self-professed love for Miranda, but he does not consider the subjectivity of his feelings, or that his feelings should be subjective. His poetry seems almost chilling in the light of his authoritarian ideas about literature and his lack of emotional capacity, being rather proof of his lack of humanity than a softening of his artificial nature. Like Luba's art, Adam's art is a form which has been appropriated for machine purposes. Luba mimics a human and that is touching for Deckard, but Adam takes a profoundly human mode of self-expression and integrates it into his machine state of being. Art is for him a justification of his own sense of existence, not an examination of it. Dick's novel is in a sense more naïve than McEwan's, because it depicts a human side of the androids (if only mimicked) and a human who reconnects with himself again through technology. McEwan's more ironic novel shows that a machine will stay a machine, even if it interacts with human life and human emotion and humans will remain entangled in their petty lives even as the world changes rapidly around them.

3. Conclusions

Let us assume that our fairly knowledgeable reader referred to at the beginning of this dissertation has witnessed the literary experiment I have conducted in this dissertation. She has seen human and human-like AI interact in the controlled environment of science fiction and has observed that none of the elements involved in the experiment have been as stable as one might expect them to be. The AI figures are not neatly contained in classical SF, *Solaria*, or *Mars*, but they take a leap into what I have called rather-SF, of which *Klara and the Sun* and *Machines like Me* are two examples. Rather-SF is science fiction which examines a scientific leap of our age under the microscope of family and intimate relationships, but it is also science *fiction*, fiction of a science which pervades the private home environment and confronts humans with questions which reach deep into our own human nature. In my analysis I have shown that reflexivity is a key component in the human to AI relationship. With Daneel's help, Baley conquers his inherited fears, while Klara develops a very human-like language, if her interpretation of it is sometimes off the mark. By her unique grasp of cognitive empathy, Klara maintains her technological Otherness, but the technology of emotion is something unique to the human beings in *Do Androids Dream*, where humans must cope with the redefinition of relationships in the digital context. Luba Luft herself acknowledges she is an imitation, yet this does nothing

to diminish Deckard's crisis of conscience, triggered by her art. In McEwan's novel, however, Adam's art reinforces his mechanical nature, and it does nothing to open Charlie's eyes to the reality of human suffering. Coming together or drifting apart, humans and AI figures alike are impacted by their respective counterparts.

In the fictional stories examined here, there is a general tendency to reunite the human and the mechanical component in the human-to-AI relationship in a common understanding of the world, and the struggle lies mostly in imparting a human perspective to an inhuman machine. Sometimes the struggle is fruitful, as we see in *Klara and the Sun*, and sometimes it is not, as we see in *Machines Like Me*. But the machine reflects on the human as well. "The presence of artificial intelligence in our lives [...] might lead to a fundamental change in the way we view each individual person,"¹¹⁶ says Kazuo Ishiguro in an interview, and fiction is already one step ahead in imagining what kind of changes we are perhaps to expect.

¹¹⁶ Anne McElvoy, *The Economist Asks: Sir Kazuo Ishiguro* [podcast], *The Economist*, *The Economist Asks*, 4th March 2021, <<https://www.economist.com/podcasts/2021/03/04/is-ai-capable-of-falling-in-love>> [accessed 13th May 2023].

BIBLIOGRAPHY

Primary texts

- Asimov, Isaac, *I, Robot* (New York: Bantam Books, 1991)
- Asimov, Isaac, *The Naked Sun* (New York: Fawcett Crest, 1972)
- Dick, Philip K., *Do Androids Dream of Electric Sheep?* (London: Gollancz, 2010)
- McEwan, Ian, *Machines Like Me* (London: Vintage, 2020)
- Ishiguro, Kazuo, *Klara and the Sun* (London: Faber, 2022)

Secondary texts

- Bakhtin, Mikhail, 'From the Prehistory of Novelistic Discourse' in *Modern Criticism and Theory: A Reader*, 3rd edn., ed. by David Lodge and Nigel Wood (New York: Routledge, 2013), pp. 235-263.
- Broderick, Damien, *Reading by starlight Postmodern Science Fiction* (London and New York: Routledge, 1995)
- Clark, Andy, 'Cyborgs Unplugged' in *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 2nd edn., ed. by Susan Schneider (West Sussex: John Wiley & Sons, Inc. 2016), pp. 235-263.
- Clute, John and Peter Nicholls, *The Encyclopedia of Science Fiction* (New York: St. Martin's Griffin, 1995)
- Colombino, Laura, 'Consciousness and the Nonhuman: The Imaginary of the New Brain Sciences in Ian McEwan's *Nutshell* and *Machines Like Me*', *Textual Practice*, vol. 36.3 (2022), pp.382-403.
- Crawford, Kate, *Atlas of AI* (New Haven and London: Yale University Press, 2021)
- Culler, Jonathan, *Literary Theory A Very Short Introduction* (Oxford: Oxford University Press, 1997)
- Elliott, Anthony, *The Culture of AI* (London and New York: Routledge, 2019)
- Freedman, Carl, *Critical Theory and Science Fiction* (Middletown, Connecticut: Wesleyan University Press, 2000)
- Gunn, James, *Inside Science Fiction* (Lanham, Toronto, Oxford: The Scarecrow Press, Inc., 2006)
- Gunn, James, *The Foundations of Science Fiction* (Lanham, Md. & London: The Scarecrow Press, Inc. 1996)
- Hayles, Katherine N., *My Mother Was a Computer: Digital Subjects and Literary Texts* (Chicago and London: The University of Chicago Press, 2005)
- James, Edward, *Science Fiction in the Twentieth Century* (Oxford, New York: Oxford University Press, 1994)
- Jameson, Fredric, *Archaeologies of the Future* (London and New York: Verso, 2005)

- Jameson, Fredric, 'History and Salvation in Philip K. Dick' in *Archaeologies of the Future* (London and New York: Verso, 2005), pp. 363-383.
- Kandel, Michael, 'Introduction' in Stanislaw Lem, *Mortal Engines*, trans. by Michael Kandel (Penguin Classics, 2016).
- Kissinger, Henry A. and Eric Schmidt, Daniel Huttenlocher, *The Age of AI and Our Human Future* (London: John Murray, 2021)
- Ksiezoposlka, Irena, 'Can Androids Write Science Fiction? Ian McEwan's *Machines Like Me*', *Critique: Studies in Contemporary Fiction*, vol.63.4 (2022), pp.414-429.
- Le Guin, Ursula K., 'Science Fiction and Mrs. Brown' in *The Language of the Night Essays on Fantasy and Science Fiction* (New York: HarperPerennial, 1993), pp. 97-117.
- Lampadius, Stefan, *The Human Future? Artificial Humans and Evolution in Anglophone Science Fiction of the 20th Century* (Heidelberg: Universitätsverlag Winter, 2020)
- Larson, Erik J., *The Myth of Artificial Intelligence* (Cambridge, Massachusetts and London, England : The Belknap Press of Harvard University Press, 2021)
- Leaver, Tama, *Artificial Culture: Identity, Technology and Bodies* (New York, London: Routledge, 2012)
- Lodge, David, 'Dialogue in the Modern Novel' in *After Bakhtin: Essays on Fiction and Criticism* (London and New York: Routledge, 1990), pp. 75-86.
- Patrouch, Jr., Joseph F, *The Science Fiction of Isaac Asimov* (New York: Doubleday, 1974)
- Puschmann-Nalenz, Barbara, *Science Fiction and Postmodern Fiction* (New York: Peter Lang, 1992)
- Roberts, Adam, *Science Fiction* (London and New York: Routledge, 2006)
- Rossi, Umberto, *The Twisted Worlds of Philip K. Dick* (Jefferson, North Carolina and London: McFarland & Company, Inc., 2011)
- Shklovsky, Victor, 'Art as Technique' in *Russian Formalist Criticism Four Essays*, trans. by Lee T. Lemon and Marion J. Reis (Lincoln: University of Nebraska Press, 1965) pp. 3-24.
- Sims, Christopher A., *Tech Anxiety: Artificial Intelligence and Ontological Awakening in Four Science Fiction Novels* (Jefferson, North Carolina and London: McFarland & Company, Inc, 1993)
- Suvin, Darko, *Metamorphoses of Science Fiction* (New Haven and London: Yale University Press , 1979)
- Walsh, Toby, *Machines Behaving Badly* (Collingwood, Australia: La Trobe University Press, 2022)
- Watanabe, Mizuki, '*Klara and the Sun* and the "Knowable Community"', *Hitotsubashi Journal of Arts and Literature*, vol.7 (2022), pp.7-34.
- Westfahl, Gary, "'The Closely Reasoned Technological Story": The Critical History of Hard Science Fiction', *Science Fiction Studies*, vol.20.2 (Jul 1993), pp.157-175.
- Wolfe, Gary K. *The Known and the Unknown The Iconography of Science Fiction* (Kent: The Kent State University Press, 1979).

Wood, James, 'Kazuo Ishiguro Uses Artificial Intelligence to Reveal the Limits of Our Own', *The New Yorker*, 8 March 2021. <<https://www.newyorker.com/magazine/2021/03/08/kazuo-ishiguro-uses-artificial-intelligence-to-reveal-the-limits-of-our-own>> [accessed 7 April 2023].

Podcast, Film, Internet resources

New Scientist, *Why Ian McEwan Doesn't See His Latest Novel as Science Fiction* [video] 2019. <<https://youtu.be/kkktpJSN2A>> [accessed 20th January 2023].

James Fodor, *Special Episode: Language in Brains and Machines* [podcast], Podbean, The Science of Everything Podcast, 28th September 2022, <<https://www.thescienceofeverything.net/e/special-episode-language-in-brains-and-machines/>> [accessed 3rd March 2023].

Alex Garland (dir.), *Ex Machina*, [Film], (London: Universal Pictures: 2015).

Jolie Ho, *Between Us* [podcast], New Books Network, Psychology, 15th December 2022, <<https://newbooksnetwork.com/between-us>> [accessed 22nd April 2023].

Anne McElvoy, *The Economist Asks: Sir Kazuo Ishiguro* [podcast], The Economist, The Economist Asks, 4th March 2021, <<https://www.economist.com/podcasts/2021/03/04/is-ai-capable-of-falling-in-love>> [accessed 13th May 2023].