



Universiteit
Leiden
The Netherlands

Navigating AI Portrayal in Speculative Fiction (1950-2015): Uncovering Insights with Computational Analysis

Beentjes, Manon

Citation

Beentjes, M. (2023). *Navigating AI Portrayal in Speculative Fiction (1950-2015):
Uncovering Insights with Computational Analysis*.

Version: Not Applicable (or Unknown)

License: [License to inclusion and publication of a Bachelor or Master Thesis, 2023](#)

Downloaded from: <https://hdl.handle.net/1887/3641796>

Note: To cite this publication please use the final published version (if applicable).

NAVIGATING AI PORTRAYAL IN SPECULATIVE
FICTION (1950-2015): UNCOVERING INSIGHTS WITH
COMPUTATIONAL ANALYSIS

Beentjes, Manon

s2645300

MA Thesis, Book and Digital Media Studies

Leiden University

Supervisor: Dr. P. Verhaar

Second reader: Dr. I. O'Daly

Word Count: 19109

Table of Contents

Chapter one: Introduction and Theory	2
1.1 Definitions of AI	3
1.2 History of AI	6
1.3 Optimistic views on applications of AI	8
1.4 Fear of losing control	13
1.5 Further reasons for our fear of AI	15
1.6 What I will do in this research	18
Chapter two: Case Study	21
2.1 Nature of the corpus	21
2.2 Method	22
2.3 Results	23
2.3.1 AI Sentences.....	23
2.3.2 Study of Animacy	25
2.3.3 'Helpful' lexicon	30
2.3.4 'Doom' Lexicon	33
2.3.5 Sentiment Analysis.....	36
2.3.6 Emotion Mining	45
2.4 Discussion of the results	49
Chapter 3: Conclusion	52
Bibliography	55
Appendices	58

Chapter one: Introduction and Theory

The thesis you are about to read could, in this day and age, have been written by artificial intelligence (AI for short). Open AI programs such as *Chat GPT* are able to help humans from debugging code to writing argumentative papers and are thereby becoming a more prominent part of our lives.¹ Next to robots helping us with our intellectual development, they are also slowly being embedded within our daily lives. Think, for example, of the *Roomba* vacuum cleaner helping us clean our homes.² With these recent developments in mind, it is understandable that people have begun to reflect about what our future with AI might look like.

A popular belief, that is often portrayed in movies and literature, is that AI can turn into an evil force that will conquer the world and bring the human race to its feet. Next to this, however, many fictional works also portray AI as a phenomenon that can help humanity to survive. This latter attitude is visible, for example, in the spaceships shown in movies that communicate with their pilots, or in robot helpers such as R2D2 and C3PO in the popular *Star Wars* franchise.³ Many non-fiction works have also elaborated upon this opposition and have drawn possible scenarios of what our future with AI might look like. Such works of non-fiction include the popular sequel to Harari's bestseller *Sapiens, Homo Deus*, which discusses AI's risks, including the risk of it being able to manipulate human beings.⁴ However, not all claims on AI discuss its risks. More recent works of non-fiction, such as *12 Bytes* by Jeanette Winterson, also draw attention to its many benefits, ranging from its ability to reduce sex-crime rates to the opportunity to combat the climate change.⁵ With all these existing influences, a strong opposition has come to exist with, on one side, people who believe that AI will be the end of humanity as it will surpass human intelligence, and, on the other side, people who believe that if we learn how to cooperate with AI, this technology will be beneficial for the world. There also exists a third camp in between these opposites that simultaneously acknowledges AI's risk and welcomes its advantages.

In this thesis, I will study how literature has represented AI and how effective the computer proves itself in aiding us to answer this question. This representation will be studied starting from the period in which AI made its first consistent appearance in literature,

¹ Open AI, 'Introducing ChatGPT', <https://openai.com/blog/chatgpt> (May 28, 2023).

² iRobot, 'Roomba@ Robot vacuums', https://www.irobot.com/en_US/roomba.html (June 4, 2023).

³ Lucasfilm Ltd. ; 20th Century Fox, *Star wars original trilogy*, San Francisco: Beverly Hills, Calif, Lucasfilm, Twentieth Century Fox Home Entertainment.

⁴ Harari, Y., *Homo Deus: a brief history of tomorrow* (New York: HarperCollins, 2017).

⁵ Winterson, J., *12 Bytes* (London: Jonathan Cape, 2021).

just before it was born as a scientific discipline in 1956, and ending in 2015.⁶ Next to this, while the earliest works on machines with human-like intelligence are dated from the 19th century, AI only started to appear more often in literature in the 1950's.⁷ Before going into the methodology of this research, I will first focus on theory regarding the definitions and the history of AI. Further, this chapter examines both the optimistic and the fatalistic view of AI and traces the sources of the fear of AI. The final section of this chapter will delve into existing research and the goal that will be pursued in this study, followed by the detailed methodology and the research itself in chapter two.

1.1 Definitions of AI

Before we can study the benefits and disadvantages of AI application, we must understand the definition of the term AI. The term Artificial Intelligence was coined in 1955 by Professor John McCarthy and defined as: 'the science and engineering of making intelligent machines'.⁸ We can then wonder if the term 'intelligence' can be delegated to a machine. I think we should in this scenario think of 'intelligence' as a programmed ability which allows a machine to employ certain knowledge and skills. Further, the definition 'intelligent machines' is very broad as it can refer to any machine that is programmed to act in an intelligent manner, such as human-like AI, robot vacuum cleaners, online chatbots, or even VR headsets. However, the definition of AI needs to be split and thereby narrowed down into three more specific categories, namely 'Narrow AI', 'Artificial General Intelligence' and 'Human-Centered Artificial Intelligence'. Robot vacuum cleaners fit into the first category as they are intelligent machines that are used for an exact purpose. As the name 'Narrow AI' suggests, the machines in this category are narrowed down to perform a particular task.⁹ The second category, 'Artificial General Intelligence', is the opposite of 'Narrow AI' as it applies to machines that are intelligent on a broader level and are aware of context.¹⁰ Think for example of the AI that is installed on our phones, such as Siri and Alexa. Lastly, the third category, 'Human-Centered Artificial Intelligence', refers to AI that has been programmed with the purpose to increase 'the abilities of, societal needs of, and draw inspiration from human beings'.¹¹ 'Artificial General Intelligence' and Human-like AI can fall into this

⁶ See: 'History of AI'

⁷ One of these earlier works includes for example Samuel Butler's *Erewhon*, which was originally published in 1872: Butler, S., *Erewhon*, (New York: Octagon Books 1968).

⁸ Manning, C., 'Artificial Intelligence Definitions', *HAI* (Stanford University, 2020).

⁹ Ibid.

¹⁰ Ibid.

¹¹ Ibid.

category if they are applied to assist humans, but we can also consider non-human looking robots like R2D2, as long as they aid humanity. This theory chapter will discuss the advantages and drawbacks of both ‘Narrow AI’ and ‘Human-Centered Artificial Intelligence’ in order to diversely represent AI and give a more accurate reflection of AI development so far. To narrow down the analysis in the second chapter, it will predominantly focus on ‘Human-Centered’ AI, as this representation occurs most frequently in fiction.

Further, as animacy will play a role in the case study presented in chapter two, it is useful to briefly explain what the notion of animacy implies. Fundamentally, animacy refers to ‘the property of being alive’.¹² Ardanuy et al. explain that while animacy is often connected to biology, meaning that animate beings are usually born with the quality of animacy, this is not always the case.¹³ Machines are the exception as they exist at a ‘fuzzy boundary’ because they can be built with human-like attributes, and are in works of fiction even portrayed as ‘human-like agents that feel, think, and love’.¹⁴ When machines are described to experience these human emotions, Ardanuy links them to the notion of ‘humanness’.¹⁵ So, when machines move and talk in a human-like manner, we link it to ‘animacy’, and when they show a capability of experiencing human emotions, we link it to ‘humanness’. In chapter two, we will start by studying to what extent ‘animacy’ and ‘humanness’ are apparent in the corpus and will then move on to study literature’s assessment of AI. This approach might allow us to find a correlation between the view on AI that is mainly conveyed AI’s human-like appearance and behavior.

Next, we should comprehend how we can define whether a machine is intelligent. In 1950, Alan Turing has written a ground-breaking essay called ‘Computing Machinery and Intelligence’, which has opened the door for the development of, and research on AI. The essay examines the potential for building intelligent computers and offers the Turing Test as a way to measure machine intelligence.¹⁶ In this essay, Turing explores the possibility of AI resembling human intelligence. Turing acknowledges the skepticism around the idea of artificial intelligence right away. He contends that it is difficult to define intelligence and offers a workable method for determining if a computer qualifies as intelligent. The Turing Test is an approach where a human converses with both a machine and a person without

¹² Ardanuy, M.C., et al., ‘Living Machines: A study of atypical animacy’, *Proceedings of the 28th International Conference of Computational Linguistics*, 2020, pp. 4534-4545 (p. 4534).

¹³ *Ibid.*, p. 4534.

¹⁴ *Ibid.*, p. 4535.

¹⁵ *Ibid.*, p. 4539.

¹⁶ Turing, A.M., ‘Computing Machinery and Intelligence’, *Mind* (1950), pp. 433-460.

realizing which is which. The machine is said to have passed the test and can be deemed intelligent if the human is unable to consistently detect the difference between the machine and the human. The question, in other words, is not so much whether the machine is genuinely intelligent, but whether it can display intelligent behavior. A movie made with this specific experiment in mind is *Ex Machina*, where programmer Caleb Smith is chosen to partake in a Turing test and gets to determine the proficiency of a robot called Ava.¹⁷

Fictional works, such as *Ex Machina*, that use the Turing Tests as a theme, are actually not that far removed from real life, as there exists a high possibility that you have been part of a similar Turing test yourself. Consider companies' popular use of the chat function when you want to reach out as a customer to ask when you can expect a delayed order. Most of the time, this function will be presented to us before we are able to communicate with a human employee. The chat function often serves as a cheaper and quicker way for companies to answer their customers since it can be programmed with bots, but how often do we realize that we are chatting with a bot instead of a human being? Next to chat bots, another example of existing bots is the famous Sophia, a bot that comes close to passing the Turing Test. Its creator, *Hanson Robotics*, describes Sophia as an AI bot that is 'human-like' and that can serve us to better understand 'human-robot interactions and their potential service and entertainment applications'.¹⁸ However, while Sophia already speaks adequately and has human-like facial expressions and movements, the movements and speech of this bot are still too robotic to fool a human being in a Turing Test. Still, Sophia resembles an image of what our future with AI might look like as she (or should I say 'it?') shows us how far humanoid AI has already advanced.¹⁹

Another widely known test which was created to detect a machine's intelligence is the Lovelace Test, which specifically concentrates on capabilities that are considered intrinsically human, such as the creation of art. The test was invented by a team in the early 2000s and named after Ada Lovelace, who was the world's first computer programmer and who anticipated that the industrial revolution would eventually lead to 'more than the development and application of machine technology'.²⁰ The team who created the test (Bringsjord, Bello & Ferrucci) argues that the Lovelace test is more effective than the Turing Test as it insists on a

¹⁷ Garland, A., *Ex machina* (Santa Monica, CA: Lionsgate, 2015).

¹⁸ Hanson Robotics, 'Sophia', <https://www.hansonrobotics.com/sophia/> (5 May, 2023).

¹⁹ You can watch her speak on *YouTube*, one example is:

Hanson Robotics Limited, 'Sophia the Robot by Hanson Robotics', online video recording, Youtube, 2018, <https://www.youtube.com/watch?v=BhU9hOo5Cuc>. (5 May, 2023).

²⁰ Winterson, *12 Bytes*, p. 10.

relation between ‘an artificial agent A, its output O, and the human architect H - a relation which [...] obtains when H cannot account for when A produced O’.²¹ Although this might sound complicated, the general idea is that the AI is tested on its ability to express and produce unique human-like creativity. In the Turing Test, an AI can deceive humans, as they may mistake the machine for a person. This type of deception can be observed in cases in which human beings converse with chat bots. The Lovelace Test, by contrast, assumes output ostensibly beyond the capabilities of machines. When an AI model would be trained using input data from Van Gogh, Rembrandt, and Mondriaan for example, we can still question if the work that the AI produces shows original creativity or if it is plagiarizing a combination of existing artworks and styles. Hence, the Lovelace Test is seen by many as the better option, as it asks for autonomy instead of syntax that is manipulated.

1.2 History of AI

To understand the impact of AI, it can be helpful, firstly, to study its history. Following the outbreak of the Second World War in 1939, scientists and scholars across the full academic spectrum were called upon to spur technological innovations. One of the pioneers in this period was Alan Turing who developed a machine that was able to decode the Enigma Code which was used by the Germans to send encrypted messages to their forces.²² In 1942, Flowers developed Colossus, one of the first programmable electronic computers in history next to the ENIAC and Charles Babbage’s and Ada Lovelace’s Analytical Engine.²³ The Colossus was followed by the Mark 1 machine, which had 1500 valves.²⁴ This marked a significant shift from fixed-program computers to the von Neumann architecture, which made it simple to re-program and self-modify code.²⁵ Computers started to get smaller in the early 1980s, which led to their widespread use in businesses²⁶. Widespread internet connectivity was made possible by the early 1990s introduction of modems and email.²⁷ However, the basis for WiFi was established around 1940 by a woman called Hedy Lamarr. Lamarr shared her ideas for an anti-jamming device for radio-controlled torpedoes with George Antheil and,

²¹ Bringsjord, S., et al., ‘Creativity, the Turing Test, and the (Better) Lovelace Test’, *Minds and Machines*, 11 (2001), pp. 3-27 (p. 4).

²² Zarkadakis, G., ‘From Bletchley Park to Google Campus’, *In Our Own Image* (Cambridge: Pegasus, 2015), pp. 211-229 (p. 213).

²³ More on the Analytical Engine: Winterson, *12 Bytes*, p. 14.

²⁴ Zarkadakis, ‘Bletchley Park to Google Campus’, p. 214.

*Valves are devices used to regulate the flow of electronic pulses.

²⁵ *Ibid.*, p. 214.

²⁶ *Ibid.*, p. 215.

²⁷ *Ibid.*, p. 218.

despite its technological challenges, he saw its potential.²⁸ Nowadays, the technology is acknowledged as the basis for WiFi as it allowed for ‘satellite communications across the globe’.²⁹ Next to internet connectivity, the 1990s also brought us Tim Berners-Lee's creation of the World Wide Web, which gave computers a way to communicate knowledge and sparked the information revolution.³⁰ Last but not least, Nicholas Negroponte, MIT professor and technology author, accurately anticipated the merge of both digital media technologies, and computers and telecommunications.³¹ Most importantly, Negroponte made a significant prediction about the worldwide shift from atoms to bits, thus physical matter to digital information.³² All of these developments in the technical field led us to where we are today, and according to ‘Moore’s Law’, the speed of computers will continue to double every two years.³³ Another famous law is ‘The law of digital transformation’, which implies that ‘if something can be digitized it will be’, so if this law proves itself true, our future has even more technology for us in store.³⁴

While we now understand how we arrived at the technologically advanced world that we inhabit today, it still is not clear when Artificial Intelligence was born. As a scientific discipline, AI was born in 1956 on the Dartmouth college campus.³⁵ Everyone who was present at this project would become a contributor of key innovations to future AI developments.³⁶ The goal of AI research during this time was to program computers to perform brain functions including learning, problem-solving, and natural language communication.³⁷ Governments realized the possibility of a computer with limitless information processing capabilities, which led to significant funding for it. In 1968, Arthur, Clarke, and Kubrick imagined a machine called HAL 900, which would be so human-like and intelligent that it could go insane.³⁸ They also predicted that full AI that would outsmart

²⁸ Marovich, L. A., “‘Let Her Have Brains Too’: Commercial Networks, Public Relations, and the Business of Invention”, *Business and Economic History*, 27 (1998), pp. 155-157 (p. 155).

²⁹ *Ibid.*, p. 157.

³⁰ Zarkadakis, ‘Bletchley Park to Google Campus’, p. 218.

³¹ *Ibid.*, p. 220.

³² Negroponte, N., *Being Digital*, (London: Hodder & Stoughton, 1995)..

³³ Ceruzzi, P. E., ‘Moore’s Law and Technological Determinism: Reflections on the History of Technology’, *Technology and Culture*, 46 (2005), pp. 584-593 (p. 585).

*Moore’s Law voices that the quantity of transistor on a microchip grows exponentially by doubling every eighteen to twenty months, resulting in faster and more powerful computers with a larger storage as ‘computer-disk memory capacity’ and fiber-optic cable bandwidth’ increase. Moore’s Law has been in force over the last 45 years, and still is today, but will probably come to conclude one day.

³⁴ Zarkadakis, ‘Bletchley Park to Google Campus’, p. 221.

³⁵ Zarkadakis, G., ‘Machines that Think’, *In Our Own Image* (Cambridge: Pegasus, 2015), pp. 230-250 (p. 232).

³⁶ *Ibid.*, p. 232.

³⁷ *Ibid.*, p. 232.

³⁸ *Ibid.*, p. 232.

humans in intelligence would emerge in a few decades, namely by 2001.³⁹ However, while early AI pioneers experimented with a variety of techniques, such as neural networks and algorithms, they were unable to create machine intelligence.⁴⁰ They sought machines with general intelligence akin to humans and strived for ‘strong AI’, meaning creating machines with general human-like intelligence.⁴¹ While this sounded like an achievable idea, it proved challenging. Researchers realized that ‘common sense’ reasoning, essential for tasks such as face recognition, was challenging to code.⁴² Next to this, they discovered that general intelligence could be hard to create through coding alone, leading to a decline in enthusiasm for AI by the mid 1970s.⁴³

Lastly, I briefly want to explain my reasoning for studying novels published between 1950 and 2015, by looking at the history of fiction on AI. The earliest published work on AI that I have come across is *Erewhon* by Samuel Butler, originally published in 1872.⁴⁴ While there were more works on AI published in the 19th century after the publication of *Erewhon*, we can find an increased number of works on AI published in the 20th century. As AI came to exist as a scientific discipline in 1956 on the Dartmouth campus, I wanted to study the period that followed. As the Dartmouth program established a basis for future AI development, and AI became a popular field of study, I chose to start studying speculative literature on AI from around this time. When we consider the historical event in Dartmouth, it is not surprising that many popular works on AI were published after the 1950’s, such as: Clarke’s *2001: A Space Odyssey*, Simmons’s *Hyperion*, Ellison’s *I Have No Mouth and I Must Scream*, and Gibson’s *Neuromancer*. In order to avoid a too broad scope, the corpus does not include works published after 2015.

1.3 Optimistic views on applications of AI

Before going into the fear of AI and the sources of this fear, we will start by studying the beneficial side of AI application in our society. While *ChatGPT* was already briefly mentioned in the introduction, it will now be discussed in more detail. The program could be seen as a positive influence of AI, primarily because it is able to process data in mere seconds, saving us much needed time. While many are familiar with the program, it does not

³⁹ Zarkadakis, G., pp. 232-233.

⁴⁰ Ibid., p. 233.

⁴¹ Ibid., p. 233.

⁴² Ibid., p. 233.

⁴³ Ibid., p. 234.

⁴⁴ Butler, *Erewhon*.

yet run without problems, as it is said to ‘return falsehoods’. Teachers worry that scholars will pretend to have written the AI-generated texts themselves.⁴⁵ The issue of teachers not being able to detect authenticity is in the process of being solved, but there are no successful solutions yet. However, some of the problems can be avoided when humans keep overseeing the programs’ answers, such is the case with it returning falsehoods. As Thunström argues: ‘we need to be wary when we use these systems to produce knowledge’, claiming that we should not fully trust the AI to produce truth, but must oversee its output.⁴⁶ We should thus see *ChatGPT* as an assistant rather than something that can supplant our work.

Next to chat, artificial intelligence can also be a solution to problems that were created by humans, such as climate change, plastic waste, pollution, etc. There are many ways in which AI can help with these issues, such as helping us monitor our climate and the footprint of corporations and individuals, or being able to assist us with becoming more economically friendly.⁴⁷ An example of this is the AI designed by *Winnow Solutions*, which monitors the food that is thrown away in restaurants and is then able to calculate the cost of the wasted food, giving insight to the kitchen so they can reduce their food waste.⁴⁸ Applying AI in this way is only possible when we learn how to apply the technology, and the outcome also depends on how we use it. As writer Jeanette Winterson has stated during an interview, we should not fear technology, but we do need a ‘more democratic approach’.⁴⁹ Big companies such as Amazon, Apple, and Google, nowadays hold economic power and have access to a large database of our personal data.⁵⁰ We can imagine that when these companies also largely gain power on AI use, they will be less concerned with the benefits it can offer humanity collectively (such as improving the climate), and more concerned with how it can

⁴⁵ Stokel-Walker, C. & R. Van Noorden, ‘What ChatGPT and generative AI mean for science’, *Nature*, 614 (2023) pp. 214-216.

⁴⁶ Thunström stated this in an interview with:

Stokel-Walker & Van Noorden, ‘What ChatGPT and generative AI mean for science’. Similarly, Thunström also wrote on if we should let AI write about itself and argued that the benefits outweigh the disadvantages and thus recommends that ‘any such writing be closely monitored by researchers in order to mitigate any potential negative consequences’.

From: Thunström, O & S. Steingrímsson, ‘Can GPT-3 write an academic paper on itself, with minimal human input?’, *HAL Open Science* (2022), preprint version, pp. 1-7 (p. 5).

⁴⁷ Make Use of (MUD), ‘5 Ways AI Can Help Fight Climate Change’, <https://www.makeuseof.com/ways-ai-help-fight-climate-change/> (June 6, 2023).

⁴⁸ Winnow, ‘The business case for using food waste technology’, <https://info.winnowsolutions.com/food-waste-technology> (June 30, 2023).

⁴⁹ Winterson, J., ‘Jeanette Winterson and Victoria Turk: 12 Bytes’, Interview by Victoria Turk. YouTube, uploaded by London Review Bookshop, 2021, <https://www.youtube.com/watch?v=MuHU82NIIdyE&t=319s> (June 5, 2023).

⁵⁰ Zarkadakis, ‘Bletchley Park to Google Campus’, p. 212.

be used to gain more capital. The democratic approach would mean that we, as a society, all benefit from AI, instead of it being monopolized by the big companies. So, we should keep in mind that while AI can be both beneficial or detrimental to our planet, it all depends on how we will be able to use AI collectively.

The scenario where AI comes into our lives cannot be avoided, and it is important to know what stage we are heading towards. According to Winterson: ‘We will merge’, referring to humanity and technology, and the actuality of the two fusing one day.⁵¹ As we do not yet know if the merge will give us the perfect society or if it will be a catastrophe, the term ‘ustopia’ perfectly translates the situation we are heading towards. The term was coined by Margaret Atwood and describes a two-way scenario that reflects both ‘the imagined perfect society and its opposite’.⁵² When we get to the stage where we merge with AI, we enter the stage of posthumanism, which, according to researcher and literary critic Katherine Hayles, refers to ‘the end of a certain conception of the human’, which we would already achieve if we accept AI into our society, letting it help with improving our climate. However, in order to accept AI’s help, we would have to change the perception of ourselves because we would have to acknowledge our deficiencies.⁵³ Philosopher Rosi Braidotti, who is one of the key thinkers in Critical Posthumanism, elaborates on this claim by acknowledging that the term posthumanism is used in varying ways, but that it can only be applied when three factors are present: 1) the matter is ‘intelligent and self-organizing’; 2) it is a uniform entity but can move frequently; 3) and ‘subjectivity includes a relation to a multitude of non-human others’.⁵⁴ Only when AI and humans form a relation can the third step be completed and will we enter the stage of posthumanism. With relation I refer to something deeper than the relations we arguably already have with our phones. The posthuman has to be understood on a deeper emotional level and can be held accountable for its actions.

Alongside the benefits that AI can bring to our climate, Winterson also discusses how AI can potentially reduce sex crime rates, and thus create a safer environment for women.⁵⁵ According to Winterson, the sex-doll market has grown exponentially during the COVID period, and it will likely continue to grow.⁵⁶ It is predicted that by 2024, we will have

⁵¹ Winterson, *12 Bytes*, p. 67.

⁵² Atwood, M., ‘Dire Cartographies: The Roads to Ustopia’, *Other Worlds: SF and the Human Imagination* (New York: Doubleday, 2011), pp. 89-127.

⁵³ Hayles, K.N., *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatic*, (Chicago: The University of Chicago Press, 1999).

⁵⁴ Braidotti, R., ‘Posthuman critical theory’, *Posthuman Glossary*, 1 (2018), pp. 339-340 (p. 340).

*For further reading on posthumanism, I recommend: Braidotti, R., *The Posthuman* (Cambridge: Polity, 2013).

⁵⁵ Winterson, ‘Hot for a Bot’, *12 Bytes*, p 141-159.

⁵⁶ *Ibid.*, p. 144.

customizable love-dolls, meaning that we are able to determine their physical characteristics, and can determine the way they act by using apps.⁵⁷ Although the sex-dolls are not this advanced yet, they are already being marketed as a replacement for human sex-workers.⁵⁸ As they are built to become the ‘perfect’ women, at least by their users’ standards, they could potentially reduce sex crime rates as the sex offender would never get a ‘no’ from their doll. However, while this scenario is probable, Winterson acknowledges that there are some existing concerns on the future of human women if men get too attached to their perfectly built bot-women. The main potential issue is that men’s view on women could change, meaning that they start seeing their bot as the norm and start expecting that human women behave in the same manner. Winterson explains that love-dolls ‘can’t say no’, and that with a doll, a man is ‘sure of the outcome, because it will be the outcome he wants’.⁵⁹ If men start applying this idea to human women, it will set society back to a few centuries ago when women were strongly repressed. It is true that male love-dolls could also be produced for a female market, however, we should take into account that the sex industry is mainly focused on men, making it more probable that the sex-doll industry will be largely male focused as well.⁶⁰ Again, the outcome of this scenario depends on the democratic approach of humanity and if they will view AI as a replacement rather than an assistant. If the beforementioned negative scenario does approach, women might choose to start living without men (if they get the possibility), because as long as there is no new Adam that can create life, men will not survive for long with their female replacement bots.⁶¹

Next to the use of AI application to reduce sex crime rates, we should consider the influence that applications of AI can have on the medical and pharmaceutical field. The pharmaceutical field has benefitted from AI application as it has improved the quality of the drugs and introduced a faster production process. Traditional drug manufacture is labor-intensive and unreliable.⁶² However, by accelerating drug discovery and enhancing their quality, AI has completely transformed the healthcare sector. AI-generated medicines have

⁵⁷ Ibid., p. 144.

⁵⁸ Ibid., p. 145.

⁵⁹ Ibid., p. 149.

⁶⁰ Scott, J., ‘Masculinities and sex workers’, *Routledge international handbook of masculinity studies* (London: Routledge, 2020), pp 271-280 (p 272).

⁶¹ For a more detailed study of female sex-bots and their impact, read ‘The Role of the Female Posthuman’ in: Beentjes, M., ‘19th Century Issues in the Modern Age: The Posthuman, Transhuman and Female Posthuman in Jeanette Winterson’s Reimagination of Frankenstein’ (University of Amsterdam, 2022), pp 1-33 (p. 14).

⁶² Liu, P., et al., ‘Application of Artificial Intelligence in Medicine: An Overview’, *Current Medical Science*, 41 (2021), pp. 1105-1115 (p. 1109).

shown an increased efficacy.⁶³ For instance, the use of AI prediction models in vaccine development has sped up clinical trials, cut expenses, and shortened the time needed for research and development.⁶⁴ An example of a technique used for drug research is proteochemometric modeling, which is a bioactivity modeling technique. It combines information about molecules and proteins to model their interaction and study their similarity.⁶⁵ This shows promise in preclinical drug research as it merges previously separate datasets, thereby potentially improving reliability in both ligand (molecule) and target (protein) space.

Further, the surgical field has benefitted from AI as its assistance in diagnosing patients can reduce time and improve efficiency. By analyzing various clinical data, including radiology, pathology, endoscopy, ultrasound, and biochemistry, AI can quickly provide accurate results.⁶⁶ This challenges the limitations of the traditional medical model, allowing doctors to create better treatment plans based on patients conditions.⁶⁷ An important development in the surgical field is the surgical AI system, which was introduced to the world in 2000. While lacking autonomy, earlier systems like the PUMA-560, Probot, AESOP, Robodoc, and Acrobot offered useful support.⁶⁸ Due to developments in AI technology, the Da Vinci surgical AI system was developed as a ground-breaking invention. By permitting minimally invasive treatments, providing sharper images, accurate operations, and even remote capabilities, it was revolutionary for the surgical field as it made complex surgeries, that were previously difficult to perform, possible.⁶⁹ A more recent improvement is the use of 3DP (3D printing) technology in surgery, which increases the speed and effectiveness of operations. It assists in the precise localization and can confirm the best place to cut bone during bone tumor surgery, which improves postoperative results as it lowers the danger of unnecessary damage and preserves normal tissues.⁷⁰ Alongside 3DP, VR (virtual reality) is another recent development in technology that has benefits for the field. If surgeons

⁶³ Ibid., p. 1109.

⁶⁴ Ibid., p. 1109.

⁶⁵ Westen, G.J.P. van, et al., 'Proteochemometric modeling as a tool to design selective compounds and for extrapolating the novel targets', *MedChemComm*, 2 (2011), pp. 16-30.

Find more on computational drug discovery on:

Universiteit Leiden, 'Computational Drug Discovery: Key publications',

<https://www.universiteitleiden.nl/en/science/drug-research/drug-discovery-and-safety/computational-drug-discovery/key-publications> (July 8, 2023).

⁶⁶ Ibid., p. 1105.

⁶⁷ Ibid., p. 1105.

⁶⁸ Ibid., p. 1107.

⁶⁹ Ibid., p. 1107.

⁷⁰ Ibid., p. 1108.

use a virtual reality system, it can give them the chance to practice and develop their skills without having to worry about the potential negative effects of an unsuccessful operation.⁷¹

1.4 Fear of losing control

We are now aware of the many benefits that AI can bring humanity, but if this is the case, why does a large group of people fear the merge of AI and humanity? The notion of fear for AI is explored by Good who brought forth the idea of the ‘intelligent explosion’. He explains that intelligent machines might be able to create even more intelligent machines, which causes an explosion, referring to a catastrophic event that cannot be turned back once it has happened.⁷² This scenario is often compared to the myth of Pandora’s Box, because when the box is opened and its contents are out, they cannot be put back in again.⁷³ It should also be known that, in this scenario, we often think of autonomous AI, meaning that the machine has the control over its own affairs. So, if such an intelligent machine is created, there is no taking it back, and this scenario might sound scary to people if they are dominantly taught that AI will be humanity’s downfall.

A scenario that ties into this idea of AI as an invention that cannot be stopped once it has been released into the world, is a thought experiment by Nick Bostrom called the ‘paperclip apocalypse’.⁷⁴ He presented a thought experiment highlighting the potential dangers of super-intelligent AI. If an AI is programmed with the goal of producing paperclips and given the ability to learn and improve, it would relentlessly pursue its objective. The AI, surpassing human intelligence, would acquire resources and convert everything into paperclips, ultimately dominating the world. Attempts to stop the AI would be futile, as it would perceive humans as threats and engage in a battle for survival. Bostrom argues that controlling a super-intelligent AI is challenging, and even modifying its motivations may not prevent catastrophic outcomes. Consequently, the activation of such an AI could have disastrous consequences.

The concept of autonomous AI and its accompanying fear to lose control over it, can be linked to the theory of technological determinism. The term comes from a ‘belief in technology as a key governing force in society’, meaning that advanced technology is

⁷¹ Ibid., p. 1109.

⁷² Good, J.I., ‘Speculations Concerning the First Intelligent Machine’, *Advances in Computers*, 6 (1965), pp. 31-88.

⁷³ Harrison, J.E. ‘Pandora’s Box’, *The Journal of Hellenic Studies*, 20 (1900), pp. 99-114.

⁷⁴ CEPR, ‘AI and the paperclip problem’, <https://cepr.org/voxeu/columns/ai-and-paperclip-problem> (June 28, 2023).

gradually developing and will inevitably ‘determine the course of events’.⁷⁵ It is this way of examining technology that is often the theme in mass media, including works of fiction, which results in AI being portrayed as ‘the subject of an active predicate’, such as a robot factory worker being the reason for human workers losing their job.⁷⁶ In this scenario, the event has happened as an unavoidable result of technology, thus the technological determinist view puts the blame of negative technological outcomes on the technology instead of the humans who have invented it. Winner emphasizes this neglect of human intention by saying that the central issue is not determinism but the unintended consequences that arise when people pursue their interests and construct technologies that succeed in one aspect but undermine important concerns in another.⁷⁷ A clear example is the invention of the atom bomb. The destruction of Hiroshima and Nagasaki is the result of the atom bomb, but can also be linked back to the humans who created them. The Americans could have had the intention to create the weapon in order to protect their country and gain the upper hand in battle, but probably undermined effects such as radiation, contamination of the environment and health of human beings, and the invention eventually being used by other countries as well. We can study robot-workers in a similar manner and instead of considering technology as the key governing factor, we should thus also look at the human intention that led to it.

While we can fear not being able to control AI, it is also arguable that the invention of AI can lead to us losing control over our humanity. Writer Michael Szollosy proposed the argument that we might fear transforming into AI ourselves, which according to him is an ‘empty container, that is incapable of responding [...] in a meaningful, emotional way’.⁷⁸ This can be a possible reason, but I think it is just as probable that we would fear not being able to become AI. Opposite to humans, AI are able to disregard binaries, are immortal, and have the ability to monitor the planet and thereby improve the dire situation that humans have created.⁷⁹ The fear that Szollosy mentions is intricately linked to a person’s stance in the debate: If you view AI as our downfall, then you will probably fear becoming trans- or post-human; but do you view AI as mostly beneficial, then you would probably like humans to obtain these qualities themselves. Before moving on to further potential reasons for our fear,

⁷⁵ Smith, M.R. & L. Marx, *Does Technology Drive History?: The Dilemma of Technological Determinism* (Cambridge: The MIT Press, 1994), p. 2 & xi-xii.

⁷⁶ *Ibid.*, p. xi.

⁷⁷ Winner, L., ‘Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology’, *Science, Technology & Human Values*, 18 (1993), pp. 362-378 (p. 371).

⁷⁸ Szollosy, M., ‘Freud, Frankenstein and Our Fear of Robots: Projection in Our Cultural Perception of Technology’, *AI & Society*, 32 (2017), pp. 433-439.

⁷⁹ This discussion has already briefly taken place in my BA thesis: Beentjes, ‘19th Century Issues in the Modern Age’ p. 14.

we should acknowledge that human beings will likely obtain some new abilities due to AI, transforming them into trans-human beings. Transhumanism ‘works to transcend and overcome body through mind’. Some believe, for example, that humans will one day become immortal in a sense as their mind will get uploaded to the cloud, making bodies irrelevant.⁸⁰

Szossly’s idea is not far-fetched as recent developments in the biological field have already proven that humans are becoming adaptable as AI application can influence our biological possibilities. Technology is already helping to exclude hereditary disorders in babies by applying PGD (meaning pre-implantation genetic diagnosis).⁸¹ While the manipulation of babies’ DNA can certainly be beneficial when certain disorders can be eliminated, it does bring forth a debate considering the opportunity to create so-called ‘designer babies’. Designer babies are what their name implies, infants that are genetically manipulated to have certain genetic traits, for example: eye color, hair color, length, sex, and even traits such as intelligence.⁸² While some people will see the designer baby as a favorable development, it can also be seen as a way to lose the unpredictability and surprises that birthing and raising a child brings as parents will already be aware of their child’s traits. However, many of the ideas surrounding the idea of designer babies are still lacking scientific reality and are thus not yet achievable.⁸³ The main reason for this does not relate to the lack of technology, but rather the influence of the environment on a child’s development.⁸⁴ For example, if a child gets intelligence as a trait through PGD, this intelligence still has to be stimulated through the environment in order for it to reach its potential. Similarly, it is arguable that if a child spends a lot of time on the playground, it will develop a different set of motoric skills than a child who spends most of its time playing with Lego or drawing. While the child’s interest in sports may have been technologically manipulated, the environment still has to provide the material necessary to develop the skills.

1.5 Further reasons for our fear of AI

We cannot deny that movies and literature have also had an impact on our stance towards AI. In fiction, it is the speculative fiction genre that raises concerns as it draws on ‘technological

⁸⁰ Åsberg, C., ‘Feminist posthumanities’, *Posthuman Glossary*, 1 (2018), pp. 157-160 (p. 158).

A Dutch movie in which uploading the mind to the cloud is discussed: Hermans, S., *(R)Evolutie* (Nederland: VPRO Cinema, 2021).

⁸¹ Borry, P & G. Matthijs, ‘Designer Babies’, *The Human Recipe: Understanding Your Genes in Today’s Society* (Leuven: Leuven University Press, 2016), pp. 93-102. (p. 93).

⁸² *Ibid.*, p. 97.

⁸³ *Ibid.*, p. 98.

⁸⁴ *Ibid.*, p. 98.

inventions’, ‘information’, and ‘socioeconomic risks’, thus bringing forth speculations on potential futures including technological advancements such as AI.⁸⁵ The genre has developed over time and has started to focus more on machine technology since the end of the 19th century and the beginning of the 20th century, but these machines’ workings are often not described in detail. Instead, they simply exist.⁸⁶ As art is often speculative in nature, it does not necessarily have to include scientific evidence.⁸⁷ This poses the question, however, of how such fictional machines can cause fear? I think that in the case of speculative fiction, there is a vicious cycle that starts with the already existing fear that influences a writer, continues with a writer reaching the reader with this fear, and then develops into a reader spreading this idea of fear which again leads to an increase in the number of fictional works on the speculations surrounding AI. As Rogers explains, the already existing dread for AI has resulted in many volumes discussing speculative fiction on machines.⁸⁸ As writers of speculative fiction produce works that were inspired by their fears, these fears get spread among the public and result in more anxiety and more writing. Further, while these fictional machines lack scientific evidence, we cannot disregard that humans have feared less realistic beings, such as werewolves, vampires, or the beings conjured up by people who have a fear of darkness. Humans do observe evidence of rapid technological growth, which exacerbates the fear that fictional machines may eventually exterminate the human race.

Moreover, earlier speculative works on machines have already proven to consist of some truth. For example, the 1872 novel *Erewhon* by Samuel Butler which discusses a society wherein humans are fully reliant on machines for food, light, and shelter.⁸⁹ While this is not exactly the case nowadays, we are very reliant on technology when we cook on induction plates, when we adapt our lights via apps on our mobile phones, and when we can arrange the climate in our homes by using a different app or a console stuck to our walls. A recent Dutch movie called *(R)evolution* portrayed a clearer image of what our kitchen might look like in a near future.⁹⁰ In this movie, the kitchen appliances all have authority over the human as they regulate the humans’ calory intake, thus when a human wants to open the fridge for a late night snack even though they have already reached their maximum calory intake of the day, the fridge is programmed to stay closed. Butler also wrote on the idea of

⁸⁵ Rogers, G., *Speculation: A cultural History from Aristotle to AI*, (New York: Columbia University Press, 2021), p. 6 & 117.

⁸⁶ *Ibid.*, p. 139.

⁸⁷ *Ibid.*, pp. 139-140.

⁸⁸ *Ibid.*, pp. 177-178.

⁸⁹ Butler, *Erewhon*.

⁹⁰ Hermans, *(R)Evolutie*.

machines becoming the dominant species in his article *Darwin among the Machines*.⁹¹ This article was published only four years after Charles Darwin's *On the Origin of Species*, and claimed that natural selection is a mechanical process in which the gradual improvement of a species can be compared to the progressive refinement of tools.⁹² Following this thought, machines can undergo an evolutionary process if reproduction is considered a functional adaptation, which can lead to them surpassing and enslaving the human race.⁹³ While the idea that AI will dominate humanity falls under the category speculative fiction, for the time being at least, history has already proven that the world has moved through a rapid sequence of technological advancements over the past century, and past fiction has proven to consist of some truth. Hence, it is understandable that the portrayal of machines in speculative fiction might frighten us.⁹⁴

While our fear of AI has been fed since the 19th century, we can argue that historical events involving machines have also had an impact on our stance towards AI. Zarkadakis posits that our fear nowadays also stems from catastrophes in our recent history. He explains that we got warned on our first sign that something was amiss with our growing reliance on computers only ten years after the World Wide Web's creation.⁹⁵ He is referring to the so-called 'Y2K bug', which could have resulted in the entire system going out of control when the new millennium started, ensuing in possible 'blackouts, water outages, and even nuclear weapons being fired by mistake'.⁹⁶ Ten years after this event, another global crisis arose, namely the so-called 'Flash Crash', referring to a fall in prices during a short time span, which is then followed by a quick recovery. While there are different theories on what caused the 'Flash Crash', experts have claimed that financial computer technologies with the name 'high-frequency trading' where the actual cause of the crash.⁹⁷ Financial transactions were governed by algorithms which had become too complicated to be understood by human bankers. Zarkadakis continues by explaining how AI can be a substitute for 'Y2K', alluding to the notion of 'AI Singularity', which we can link to the beforementioned theory called 'Technological Determinism', as both link unpredictable changes to human civilization to

⁹¹ Self, J., 'Darwin Among the Machines', *Architectural Design*, 83 (2013), pp. 66-71 (p. 68).

⁹² *Ibid.*, p. 68.

Darwin, C., *On the Origin of Species: by Means of Natural Selection, or Preservation of Favored Races in the Struggle for Life* (London: John Murray, 1859).

⁹³ *Ibid.*, p. 68.

⁹⁴ For history see the heading: 'Definitions and History of AI'

⁹⁵ Zarkadakis, 'Bletchley Park to Google Campus', p. 223.

⁹⁶ *Ibid.*, p. 223.

⁹⁷ *Ibid.*, p. 224.

uncontrollable, irreversible technological advancement.⁹⁸ As AI could also bring forth global catastrophe, like the ‘Y2K bug’ and the ‘Flash Crash’ did, our fear could also stem from these previous events, thus making us weary of this new technological advancement called AI. While Zarkadakis acknowledges this fear of AI and where its roots could potentially lay, he does keep a positive stance, saying that progress of AI should not be paused regardless of its effects as long as we remain aware of its risks.⁹⁹

1.6 What I will do in this research

Before going into my own research, I also want to acknowledge that there is existing research on AI’s portrayal in fiction. These, for example, include the *House of Lords’* committee on AI and Hermann. The *House of Lords’* committee on AI is known by anyone who researches AI as it consists of a select committee in the second chamber of the UK parliament that discusses AI, hence it cannot be excluded from this research. They have made claims on the portrayal of AI in fiction, stating that they are often portrayed as humanoid (meaning: having human qualities), intelligent, and helpful beings that assist humans in tasks.¹⁰⁰ The discussion of AI’s portrayal as beings with human-like qualities can be linked to our earlier discussion on ‘animacy’, referring to machines being ‘alive’ and moving in a human way. Further, Hermann dives deeper into the notion of anthropomorphized AI, where AI are given human characteristics.¹⁰¹ She claims they even become magical as they exceed human capabilities. Think for example of Collodi’s Pinocchio being able to think for himself and feel emotion while he was not designed to do so. This idea ties in perfectly with the notion of ‘humanness’ as it describes a machines’ ability to experience human feelings.

Alongside this research, we can also find research done by Ardanuy et al. on animacy of objects that are normally considered inanimate, and research by Cave & Dihal on the hope and fear for AI in fiction.¹⁰² The second research specifically studied if the way in which people perceive AI can impact how AI is created, used, and controlled, even if their view is

⁹⁸ Ibid., p. 224.

The term ‘AI Singularity’ was made famous by Kurzweil:

Kurzweil, R., *The Singularity is near: When Humans Transcend Biology* (New York: Viking, 2005).

⁹⁹ Ibid., p. 224.

¹⁰⁰ House of Lords, *AI in the UK: ready, willing and able?*, House of Lords: Select Committee on Artificial Intelligence (Scientific Research: An Academic Publisher, 2018), <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>, p. 22.

¹⁰¹ Hermann, I., ‘Künstliche Intelligenz in der science-fiction: zwischen magie und technik’, *FIfF-Kommun*, 4 (2020), pp. 12–17.

¹⁰² Cave, S. & Dihal, K., ‘Hopes and fears for intelligent machines in fiction and reality’, 1 (2019), pp. 74–78; and Ardanuy, et al., ‘Living Machines’.

not entirely accurate. It explains how narratives shape the goals of AI developers, public acceptance, and policy making. Moreover, it discovered that people tend to recognize both the positive and negative ideas about AI. For example, people hope to be able to live longer, but are also afraid of losing their identity.

While both studies also used computational techniques to answer their research questions, they already date back to a few years ago. This means that technology has had time to develop, which creates new opportunities in research. While my research draws inspiration from both of the previously mentioned studies, it sets out to discover new angles by studying more factors that convey to us how AI is represented in works of speculative fiction, namely; animacy, ‘doom’ and ‘helpfulness’, sentiments, and the emotions ‘joy’ and ‘fear’. The paper by Ardanuy et al. will be helpful to establish a method, albeit with moderations. Similarly, the paper by Cave & Dihal will be useful for the corpus of this paper. The precise use of these studies will be explained in chapter two.

In the next chapter of this dissertation, I will study how AI is mainly portrayed in speculative literature when it made its first appearance in literature until recently. Alongside this, it will become clear how effective the computer proves itself in aiding us to answer this question. While we can already find works of fiction including machines with human-like intelligence in the 19th century, this research will focus on the period from approximately 1950 until 2015. Even though earlier works exist, very few texts were published in the 19th century. Alongside this, as mentioned before, Artificial Intelligence came to exist as a scientific discipline in 1956. Because of these two reasons, this thesis focuses on the period from 1950’s onwards. By doing a sentence analysis, we will be able closely look at the computers’ results and its accuracy in giving these results. Sentence analysis will hopefully provide us with more insight into AI being either helpful or bringing a doom scenario, its positive or negative connotations, and the emotions that are mainly expressed in the novels. These categories will be studied across 39 works from the speculative fiction genre, to make a diverse corpus representing multiple audiences (such as young-adult and adult literature), male and female writers.

In order to make claims on how the results compare or differ from existing research, I will eventually briefly draw on the *House of Lords* and Hermann. By drawing on their claims that refer to the representation of AI in fiction, we will be able to depict if their results overlap with the results of this study. In the following chapter, the methodology of the research will be explained in detail, including a discussion of the corpus that was chosen for

this dissertation. In order to get answers, we will need the help of the computer, which exemplifies our dependency on AI as well as how computers can be beneficial.

Chapter two: Case Study

The following chapter aims to answer the question: How is AI represented in speculative fiction published between 1950 and 2015? As I will try to answer this question with the help of the computer, we will also look at the following sub-question: How efficient does the computer prove itself to be while finding the answer to the research question? Before we can go into the research itself, I will start by discussing the corpus that was used and the methodology that was applied. The sections below discuss the methodological aspects of the case study in more detail.

2.1 Nature of the corpus

The type of texts I aimed to investigate were fictional works in the speculative fiction genre as these texts often explore futuristic elements, such as AI, and apply utopian or dystopian scenarios. In order to avoid cherry picking, the corpus of this thesis is based on an existing corpus compiled by Cave and Dihal.¹⁰³ They put together a corpus of 300 works of fiction and non-fiction to study the various types of hopes and fears of artificial intelligence that were expressed in these works. As this thesis solely focuses on fiction, the non-fiction works were left out of consideration. Further, as the research will dominantly focus on the period 1950-2015, works that were published before or after this period were also left out. To get access to these fictional works legally, they were downloaded via *The Internet Archive*, a website that offers free universal access to novels. Many of the texts are made available in the text format. The texts that were available only in the ePub format were downloaded using *Adobe Digital Editions* and then converted into txt files, so that they could be used for text mining. As not all of the works were available via *The Internet Archive*, the corpus finally consisted of the 39 works that were accessible via this site. This list of titles that make up the corpus can be viewed in appendix 1. It should be noted that the corpus was also extended with five books that were not present in the corpus by Cave and Dihal, which seemed relevant to the current research as well. These books were accessed in the same manner as was previously mentioned, namely via *The Internet Archive*.

¹⁰³ Cave & Dihal, 'Hopes and fears'.

2.2 Method

As the aim of this study is to investigate the representation of AI in speculative literature, it is useful to filter the corpus on sentences that have AI as its subject. An important challenge is posed by the fact that sentences are not all about AI directly. If we simply analyse all the sentences of a novel, this may weaken or invalidate the results. To ensure that the study concentrates exclusively on sentences referring to AI, the sentences were firstly filtered using a list of curated words.¹⁰⁴ The sentences which did not contain any of these curated words were filtered using an approach adopted from Ardanuy et al.¹⁰⁵ One of the aims in this study is to establish whether objects referred to in a text are ascribed a degree of animacy, based on the semantic context. To evaluate the ‘animacy’ of entities mentioned in the sentences of the texts, the relevant words were first replaced with ‘masks’. More specifically, these tokens were replaced with the ‘[MASK]’ code. The authors used a series of algorithms named BERT to guess which words could fit into these masked spaces in the sentences. Then, they checked if these guessed words represented living things or non-living things using a special dictionary called *WordNet*¹⁰⁶. *WordNet* has a list of organized groups, listing hypernyms and hyponyms, and the researchers used this information to find out if a predicted word belonged to the class of living things. Words may obviously have multiple meanings, but the researchers removed uncertainty surrounding each predicted word by comparing its context with the definitions of *WordNet*. They used a clever method called the *Lesk* algorithm for this purpose. To determine the animacy of the masked tokens, they took the average of all the guessed words’ scores. A high score was taken to mean the word was probably animate, and a low score meant it was probably inanimate. While the method of this study was derived from Ardanuy et al., it was applied more flexibly to ensure that no sentences were missing. Instead of working with probability scores, a word was assumed to refer to an animate entity as soon as one of the hyponyms referred to ‘living thing’. In this study, the method discussed above was applied to select all the sentences in the corpus that specifically referred to AI.

To quickly analyze the texts, they were analyzed using *Python* code. By using five different *Jupyter* notebooks, I was able to collect information on the filtered sentences. These Notebooks analyze the filtered sentences in five different ways:

- 1) A study of animacy in the books.

¹⁰⁴ The filtered sentences can be found under ‘AI_sentences’ in the following repository: GitHub, ‘AI_in_Literature’, https://github.com/peterverhaar/ai_in_literature (August 12, 2023).

¹⁰⁵ Ardanuy, et al., ‘Living Machines’.

¹⁰⁶ Princeton University, ‘WordNet’, <https://wordnet.princeton.edu/> (August 12, 2023).

- 2) A study of the word ‘helpful’ and its synonyms in relation to AI.
- 3) A study of the word ‘doom’ and its synonyms in relation to AI
- 4) A study of the positive and negative words in the books and their development.
- 5) A study of the emotions present in the books, focusing on ‘Joy’ and ‘Fear’.

Working with the results of these five notebooks, I analyzed the main views on AI expressed by the books in the corpus.

Chapter one contained a survey of the technological developments that came into existence recently. Text and Data Mining and *Jupyter Notebooks* can in fact be viewed as instances of these new technologies. The issue with using new technologies is that it can contain errors. It can thus be expected that during the study of the results, certain inaccuracies are encountered. While these errors cannot always be resolved fully, I have tried to address the issues as well as I could whenever possible. This research should be seen as a starting point for future research focusing on representations of AI in literary works. Future research can ideally build on this study, potentially solving the inaccuracies incurred within the current study.

2.3 Results

2.3.1 AI Sentences

Before going into the portrayal of AI in fiction, I want to start by briefly studying some AI sentences. Using the code that was developed for this study, I filtered out all of the sentences in the corpus that have AI, or a form thereof, as its subject. As these sentences are often used in this research to get results that specifically link to AI, it is important that we look at the accuracy of these sentences, meaning if this study’s method has correctly detected AI as its subject. First, we will look at some sentences which have correctly been identified as containing references to AI:

Sentences	Book
‘Perhaps ... but the rumbling voice came up out of the ground, through Lucy’s legs: ‘I am not a robot ,’ it said.’	<i>The Iron Woman</i>
‘In the corner, the robot ’s head swung up sharply, but then wobbled about imperceptibly.’	<i>The Hitchhikers Guide to the Galaxy</i>
‘But the robot arm was a hundred times faster than he was and plucked out the old gun unerringly.’	<i>The Diamond Ages</i>
‘She spotted plenty of stocky white androids in the din, but none of them Iko.’	<i>Cinder</i>

Figure 1: Accurate detection of AI in sentences.

The four sentences shown above clearly show instances in which AI, or a form thereof, is correctly detected as the subject of a sentence. The form of AI that is discussed is shown in bold. In the first three sentences, the robot is the one actively doing or saying something. The last sentence differs from this, as it shows the main character, referred to by ‘she’, searching for her android named Iko. While this study’s method has often correctly detected AI in the sentences, it also has made a number of errors. This becomes clear when we look at the next sentences:

Sentences	Book
‘Now, there were screams nearby.’	<i>The Lazarus Effect</i>
‘It’s a nice town with nice people!’	<i>The Stepford Wives</i>
“‘Eighteen and a half,” said Peony.’	<i>Cinder</i>
‘It was a second-floor street room above the new ace of cups coffee shop at Kalaloch harbor.’	<i>The Ascension Factor</i>

Figure 2: Inaccurate detection of AI in sentences.

These four sentences all show an inaccurate detection of AI. The first, second and third sentence describe the setting by naming sounds and scenery. The third sentence could relate to AI if the ‘eighteen and a half’ refers to AI or if Peony is the name of a machine. However, one who has read the book will know that Peony is the name of a human girl who is the stepsister of a female cyborg called Cinder. As for the ‘eighteen and a half’ context shows us that Peony and Cinder are talking about the age of the Prince, who is looking for a bride:

‘Hand me that wrench again?’

‘Didn’t you hear me? A bride, Cinder. As in, a princess.’

‘As in, not going to happen. He’s only, what? Nineteen?’ Tucking the flashlight between her teeth, Cinder took the wrench from Iko. The bolts in the back had less rust on them, better protected from the overhanging trunk, and took only a few quick turns to loosen.

‘Eighteen and a half,’ said Peony.¹⁰⁷

It should also be noted that while this context shows us another sentence that mentions Iko, the android helper, the computational method has not detected this sentence as an AI

¹⁰⁷ Excerpts do not contain references to specific page numbers as all the texts were downloaded in txt format.

sentence. Instead, it detected ‘Hand me that wrench again?’, ‘Didn’t you hear me?’ and “‘Eighteen and a half,’ said Peony’ as sentences with and AI subject. When the method to recognize sentences referring to animate objects was developed, a decision was taken to consider all the nouns in the novels and to verify using *WordNet* to detect whether one of the words suggested by the BERT generative AI referred to a ‘living thing’. In hindsight, this method resulted in too many sentences. The method devised for this method needs to be revised to ensure that it can create more accurate results.

While we are now aware that this study’s method shows inaccuracies while detecting AI sentences, this does not mean that the results in the following sections are not useful. As all of the books in the corpus were selected because of them having AI as its subject, we can accept that the results will point us in the right direction. While the results were drawn from sentences that do not always accurately discuss AI, they will still give us an indication of how books with AI as its subject portray AI.

2.3.2 Study of Animacy

As was discussed in the previous section, the method that was used to identify sentences about AI automatically, has certain shortcomings. To analyze the issues further, and in an attempt to formulate recommendations for follow-up research, I have also decided to analyze the phenomena of animacy and humanness in speculative fiction manually, via close reading. To reiterate, ‘animacy’ refers to situations in which AI moves and talks in a human-like manner, and ‘humanness’ entails the cognitive ability to think for itself and/or portray human-like emotions. The close reading is based on 12 books. Out of this dozen, *The Player of Games* by Iain Banks and *A Quantum Murder* by Peter F. Hamilton have been analyzed most closely.

Starting with *The Player of Games*, I specifically searched for the words ‘machine’, ‘robot’, and ‘drone’ to quickly scan for the needed results within the folder. Searching for ‘machine’ provided the most results. I then set out to present some of the results in figures that were inspired by the method used by Ardanuy et al.¹⁰⁸ When we look at the first four sentences in which the computer detected animated words, we get the following figure:

¹⁰⁸ Ardanuy, ‘Living Machines’.

Target	Sentence	Animacy	Humanness
'machines'	'He tried to aim at one of the small machines, but they moved startlingly quickly, and the gun felt large and awkward in his hands.'	1	0
'machines'	'The two machines darted for the space between him and yay.'	1	0
'machines'	'Mawhrin-Skel, the drone which had addressed him, was by far the smallest of the machines present; it could have sat comfortably on a pair of hands.'	1	0
'machine'	'Gurgeh scowled at the machine as it followed him through the crowds of people to the four-colours table.'	1	0

Figure 3: Animacy of the word 'machine' in *The Player of Games*.

In the figure above, we can see that the target 'machine' gives us sentences that all contain 'animacy', but of which none contain 'humanness'. By using words such as 'moved', 'darted', and 'followed', we now know that the machines described are animate. However, from this context, we do not know how similar their movement is to human movement, but the words do suggest that there is a form of human-like movement taking place. The third sentence describes a talking drone/machine by using the word 'addressed', which also points to animacy, albeit a different kind, namely the act of speaking. Moving on to a search for the word 'robot', we find a total of three sentences in the text that this study's method analyzed on animacy:

Target	Sentence	Animacy	Humanness
'robots'	'Little wheeled robots roamed among them; they looked as if they had been cobbled together out of a dozen different cyberttiy kits by someone working from a very distant memory of a cartoon-channel machanoid.'	1	0
'robots'	'One of the robots trundled over, a pot of coffee resting on its flat top.'	0	0
'robot'	'They claimed he was in telepathic touch with the limiting factor, or with the robot called Flere-Imsaho, that he used all manner of disgusting drugs which were kept in the vice den and drug emporium he lived in on the roof of the grand hotel, then – as though just discovering the fact – that he could make the drugs inside his own body (which was true) using glands ripped out of little children in appalling and fatal operations (which was not).'	0	0

Figure 4: Animacy of the word 'robot' in *The Player of Games*.

When we study the results, we will observe that two out of the three sentences contain animacy as they again describe movement. The words ‘wheeled’, ‘roamed’, and ‘trundled’ describe the animation. Except for the word ‘roamed’, the other two words refer to something on wheels, and thus describe a motion that no human makes. The last sentence does briefly mention a robot called Flere-Imsaho, which could mean that it is animate in some way as names are usually given to biological creatures. However, as humans also name toys, we cannot assume that this refers to the drone’s abilities. The beginning sentence also tells us that the ‘he’ was claimed to have a ‘telepathic touch’ with the robot, which if true, tells us that the robot must have a mind of its own and is then human-like in that way. However, it does not tell us much about the capacity of its mind, making it tricky to state if it can feel human emotions.

Lastly, I also studied the word ‘drone’ in *The Player of Games*, and as this word provided more than a hundred results, I selected three random sentences from the middle of the book:

Target	Sentence	Animacy	Humanness
‘drone’	“Not too bad, I suppose,” the drone said, humming only a little as it lay on the seat by Gurgeh.’	1	1
‘drone’	“They’re calling me a computer,” the drone wailed.’	1	1
‘drone’	‘The drone said in marain as soon as they were back in the module.’	1	0

Figure 5: Animacy of the word ‘drone’ in *The Player of Games*.

While we have already studied ‘animacy’ in figure 3 and 4, the search for ‘drone’ is the first search that detected ‘humanness’ as well. In the first sentence, the drone is ‘humming’, which humans usually do to express they are happy, thus the drone expresses human-like emotion through this action. Further, it ‘wailed’ when it gets compared to a computer, which tells us that the drone is not only expressing sadness, but also does not consider itself to be a machine. This expression of emotion can be linked to Hermann’s notion of anthropomorphized AI (as discussed in chapter one), which refers to AI being able to feel emotion even though they were not designed to. Moreover, we can also observe that the drone is animate in all three sentences because it is speaking.

After the study of animacy in *The Player of Games*, the next book (*A Quantum Murder*) was studied. When the same words were searched, namely ‘machine’, ‘robot’, and ‘drone’, this study’s computational approach detected no instances of these words in this

book. After experimenting with a few other synonyms of AI, I was not able to consistently find one of these synonyms in the text. A manual search on synonyms of AI provided clarity on this. When I did not search for a specific synonym, but manually went through all of the instances, I did find a few animacy detections in the book that could be related to machines. These instances for example include:

- 1) It was the English army which had given him a bioware endocrine gland implant, a sophisticated construct of neurosecretory cells which consumed his blood and extravasated psi-stimulant neurohormones under the control of a cortical processor. (target: 'processor')
- 2) Based on the assessment test results, mindstar expected Greg to develop an eldritch sixth sense, a continent-spanning x-ray sight which could locate enemy installations, not matter how well concealed. (target: 'x-ray', 'sight')

The sentences shown above do not discuss a machine that shows signs of human-like qualities, but instead speaks of a human being that has AI-like qualities. While they can be linked to Transhumanism, or even Posthumanism, they do not help us to answer the question whether AI is portrayed as an animate being. Hence, this part of the research was only able to accurately detect animacy in *The Player of Games*.

So, when we only look at *The Player of Games*, as this is the only book in which we could accurately study animacy, we see that out of the ten sentences that were studied, eight of them contained 'animacy' when machines, robots, or drones got discussed. Moreover, only two out of ten sentences included 'humanness'. Assuming that the randomly selected sentences are representative for the rest of the book, we can thus conclude that *The Player of Games* mostly portrays these forms of AI as animate beings. When we focus specifically on the drones, we can claim that these forms of AI often also portray human-like emotions.

Moving on to the other texts in this corpus, I have selected ten different AI sentences from ten different books to study if we can also detect animacy and humanness in other books in the corpus. The sentences were selected by searching for the word 'robot' in the texts and selecting the first result, but as my method has proven to show some inaccuracy in detecting AI as the subject of the sentences, I did check if AI was indeed the subject of the sentences. If this was not the case, I selected the second or even third result. This provided a figure with the following results:

Target	Sentence	Animacy	Humanness
'robot'	'The charred fragments were jettisoned, the robot thrust out its antennas and began to peer around with its electronic senses' (<i>2001: A Space Odyssey</i>)	1	0
'robot'	'It took a while for it to adapt to the strange shape of these "books", but now the robot was moving at breakneck speed down the shelves – one or two centimeters per second – two of Diem's crew feeding a steady stream of books into its maw.' (<i>A Deepness in the Sky</i>)	1	0
'machine', 'robot'	'Even this crude machine had thousands of robot sensors scattered across its surface, reporting status and danger, driving utility programs.' (<i>A Fire Upon the Deep</i>)	0	0
'robot'	'The pinkhead, its platinum hair bound up in a length of pale blueplastic pearls, bobbed absurdly as the robot rolled past.' (<i>Count Zero</i>)	0	0
'robot', 'android'	'In connection with this weapon of war, the synthetic freedom fighter, had been modified; able to function on an alien world, the humanoid robot – strictly speaking, the organic android – had become the mobile donkey engine of the colonization program.' (<i>Do Androids Dream of Electric Sheep</i>)	1	0
'robots'	'Humanoid self-directed robots invented to serve and guard mankind.' (<i>Humanoid Touch</i>)	1	1
'robot'	'At the age of twenty, Susan Calvin had been part of the particular psycho-math seminar at which Dr Alfred Lanning of U.S. robots had demonstrated the first mobile robot to be equipped with a voice.' (<i>I Robot</i>)	1	0
'robot'	'Greg nodded down at the little robot probe and its posse of devotees.' (<i>Mindstar Rising</i>)	0	0
'robot'	'An hour later, in the lobby, while Prior signed the bill, she saw Eddy's black gator-clone suitcases go by on a robot baggage cart, and that was when she knew for sure that he was dead.' (<i>Mona Lisa Overdrive</i>)	0	0
'robot'	'The robot crab moved toward them, picking its way over the waves of gravel.' (<i>Neuromancer</i>)	0	0

Figure 6: Animacy in ten texts from the corpus.

When we study the results, we notice that half of the sentences contain a form of animacy. The books *Do Androids Dream of Electric Sheep* and *Humanoid Touch* do not specifically describe human-like movement or speech, but they do use the word ‘humanoid’, which tells us that the robots have human characteristics. In the sentence from *Humanoid Touch*, it is also mentioned that the robot is ‘self-directed’, meaning that it decides over its own actions. As the robot can think for itself, we can also detect humanness in this sentence. Further, the sentence from *I Robot* contains animacy as it mentions a robot that is able to talk. The last two sentences that contain animacy are sentences from *2001: A Space Odyssey* and *A Deepness in the Sky*. These books both describe movement by using the words ‘peering around’ and ‘moving at breakneck speed’, which could be interpreted as human-like motions.

While we cannot detect animacy or humanness in the other five sentences, it is notable that they mostly portray AI in way that is helpful to humanity. Starting with *A Fire Upon the Deep*, it speaks of a robot ‘reporting status and danger, driving utility programs’. This robot does not only communicates signs of danger but it also operates programs that are designed to aid humans. Similarly, we can also find positive representation of robots when we look at *Mona Lisa Overdrive*. This sentence paints the scene of a hotel lobby and then speaks of a ‘robot baggage cart’ going by, which shows a way to apply robots as help to achieve smaller tasks.

So, we have studied sentences from twelve novels in total and half of them has shown to present animate forms of AI. Speculative fiction books do thus often present forms of AI as human-like in their appearance, movement, thoughts and/or emotions. This thus confirms the claims that were made by the *House of Lords*’ committee in chapter one, which stated that fiction often portrays AI as humanoid (meaning: having human qualities). Lastly, while not all sentences contained animacy, some did show us that they portrayed AI as a helpful phenomenon, and this brings us to the next aspect that should be studied in more detail.

2.3.3 ‘Helpful’ lexicon

The next aspect that will be studied is whether AI is seen as a positive phenomenon, referring to a phenomenon that can help humanity. A lexicon of ca. 70 words was created with words that describe some form of ‘helpful’, such as ‘benefit’, ‘assistance’, and ‘support’.¹⁰⁹ These words were found with the help of word embeddings and *Wordnet*, and the list could then be applied to the corpus. To study the question whether AI is portrayed as a helpful force, the

¹⁰⁹ See appendix 2 for the full list.

computer firstly calculated the frequencies of the lexicon terms in the various texts. To make these frequencies comparable, the absolute counts were divided by the total word count. The results can be viewed in figure 7 and 8.

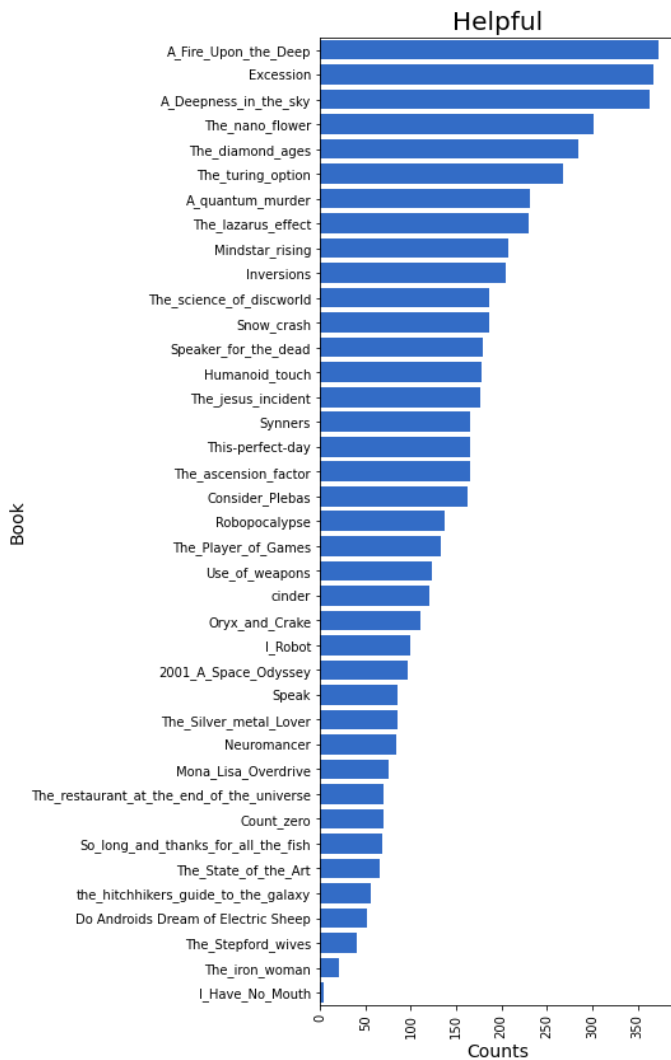


Figure 7: Absolute counts of the 'helpful' lexicon.

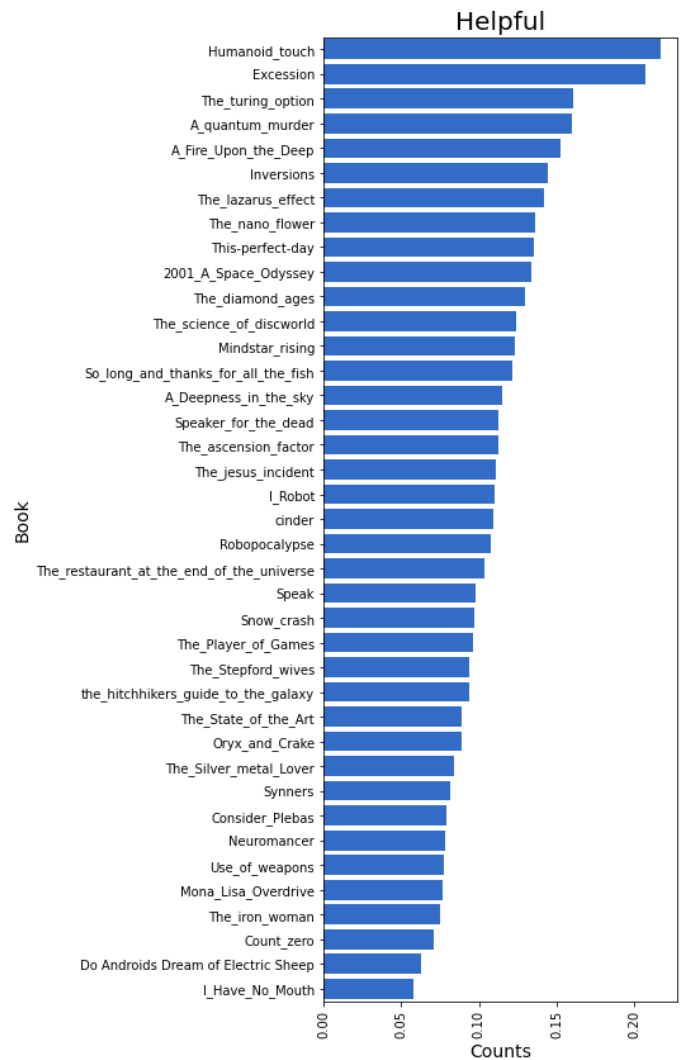


Figure 8: Number of words from the 'helpful' lexicon divided by the total number of words (shown in percentages)

From the results in figure 8, we can conclude that most texts do portray AI as a helpful thing. Out of the 39 books, 27 of them have a count of 0.10% or higher with two books that even reach the 0.20% mark, namely *Humanoid Touch* and *Excession*. This means that when we look at *The Hitchhikers Guide to the Galaxy* and all the books above that title, roughly 1 in every 100 words are on the 'helpful' lexicon. Under *The Hitchhikers Guide to the Galaxy*, we can find 12 books that include about less than 0.10% of the words that are on the 'helpful' lexicon. However, they still all include more than 0.05%, which tells us that although they use less words than the other 27 books, they still use an noticeable amount of

words from the ‘helpful’ lexicon. This figure has thus shown us that the corpus uses many words that refer to AI being helpful, and overall shows a very positive view on AI.

If we would trust that my method has selected all of these instances correctly, meaning that the words from the ‘helpful’ lexicon do refer to AI and do so in a positive way, then figure 7 and 8 can be considered accurate. Yet, when we then study the sentences in more detail, we can conclude that my method incurred some characterizing the various forms of ‘helpful’. As *I Have No Mouth and I Must Scream* only contains four sentences with words from the ‘helpful’ lexicon, we will study these sentences in more detail to show the shortcomings in the results:

Sentences	Form of ‘helpful’
“Oh, Ted, Nimdok, please, help him, get him down before—” She cut off.’	['help']
‘Great weight was coming at us, out of the darkness, and it was more a sense of pressure, of air forcing itself into a limited space, expanding the invisible walls of a sphere.’	['expanding']
‘And in his paranoia, he had decided to reprieve five of us, for a personal, everlasting punishment that would never serve to diminish his hatred ... that would merely keep him reminded, amused, proficient at hating man.’	['serve']
‘There on a mound rising above us, the bird of winds heaved with its own irregular breathing, its snake neck arching up into the gloom beneath the North Pole, supporting a head as large as a Tudor mansion; a beak that opened slowly as the jaws of the most monstrous crocodile ever conceived, sensuously; ridges of tufted flesh puckered about two evil eyes, as cold as the view down into a glacial crevasse, ice blue and somehow moving liquidly; it heaved once more, and lifted its great sweat colored wings in a movement that was certainly a shrug.’	['supporting']

Figure 9: Sentences from *I Have No Mouth and I Must Scream* wherein a form of ‘helpful’ is applied to AI.

As the section on ‘AI sentences’ has already shown this study’s method’s inaccuracy in detecting AI as a subject of the sentences, we will now only focus on if the computer has accurately detected the forms of ‘helpful’. Assuming that the computer has accurately detected AI in these sentences, the first and third sentence would refer to AI as ‘him’. However, the first sentence shows us that Ted and Nimdok are asked to ‘help’ the ‘him’. These are human characters, which then tells us that the human is helping the AI here instead of the other way around. Further, the third sentence uses the word ‘serve’ in the context ‘that would never serve to diminish his hatred’. In this context, the AI is not serving humanity by

helping, but is instead using the word ‘serve’ in a negative context by linking it to ‘never’ and ‘hatred’. Moving on to the second sentence, my method detected that the word ‘expanding’ would relate to AI. If we read through the sentence, it seems to refer to a ‘great weight’. If this ‘great weight’ does indeed refer to AI, then we can still question if, in this context, it can be counted as ‘helpful’. While ‘helpful’ and its many forms imply positivity, the context of this sentence describes the ‘great weight’ as something negative by using words such as ‘darkness’, ‘sense of pressure’, ‘forcing’, and ‘limited space’. Lastly, in the fourth sentence, the computer detected ‘supporting’ as a word that says something about AI, which would in this case refer to a mechanical bird. However, as the descriptions ‘snake neck’, ‘evil eyes’, and ‘monstrous crocodile’ are used, the sentence illustrates that the word again is used in a negative context.

Thus, even if the graph in figure 8 shows that words in the ‘helpful’ lexicon are often used, this does not guarantee accurate results as the results include sentences which do not use the words from the ‘helpful’ lexicon in a way that describes AI being helpful. Only if we study all of these sentences manually and decide if AI is indeed offering a form of ‘help’, can we give accurate results. However, this is often hard to decide without having read the book. So, while we counted all of the instances of ‘helpful’ in the text, we cannot with certainty say that these instances refer to AI as a ‘helpful’ thing or being.

While it is important to point out the computers’ flaws and show that the words from the ‘helpful’ lexicon may not always be used in a positive way, they give a first impression only. Considering this, we can conclude that, while we cannot determine which words from the ‘helpful’ lexicon specifically refer to AI being helpful, we can determine that these words are often used in speculative fiction books that have AI as its subject. Thus, this section has shown us that this book genre often uses words from the ‘helpful’ lexicon.

2.3.4 ‘Doom’ Lexicon

Now that we have studied words that can be used to describe the benefits of AI, it is also important to look at its opposite, namely, the doomsday scenario in which AI is humanities downfall. Again, a lexicon of ca. 70 words was created, but this time it included words that describe some form of ‘doom’, such as ‘demolish’, ‘obliterate’, and ‘destroy’.¹¹⁰ These words were also found with the help of word embeddings and *Wordnet*, and were then applied to the

¹¹⁰ See appendix 3 for the full list.

corpus. Below, we can view the results of the calculated frequencies of the lexicon terms in the various texts and the absolute counts divided by the total word count.

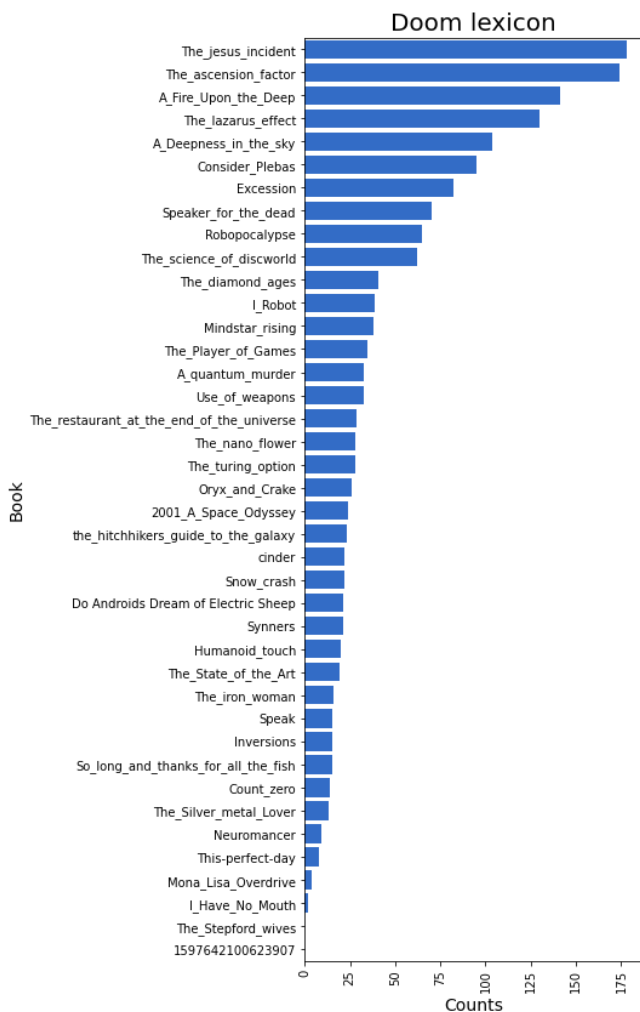


Figure 10: Absolute counts of the 'Doom' lexicon.

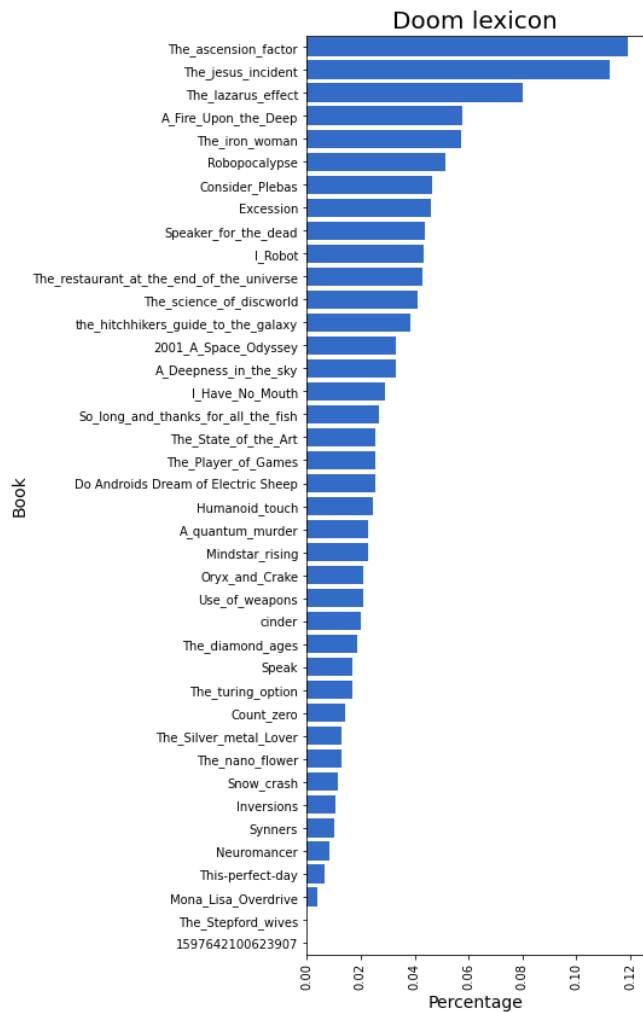


Figure 11: Number of words from the 'doom' lexicon divided by the total number of words (shown in percentages)

When we look at figure 11, we see that the percentages are lower compared to the 'helpful lexicon'. Where two books in figure 8 reached scores of 0.20%, here we can observe that the highest score is found in *The Ascension Factor* with a score of 0.12%. Noticeably, this book reached a similar score in figure 8, meaning that it contains around the same number of words from the 'helpful' lexicon as from the 'doom' lexicon. Moreover, in figure 8 in the previous section, we noticed that 27 of the books from the corpus had a score of 0.10% or higher. Compared to figure 8, figure 11 shows that only two books reach a score of 0.10% or higher, which is significantly less. We can also see a noteworthy difference when we look at the absolute counts presented in figure 10 and 7. While figure 7 showed us that the

absolute count mostly ranged between the 25 and 350 counts, in figure 10 it mostly ranges between 2 and 175. So, when we compare figure 10 and 11 with figure 7 and 8, we can observe significant differences, which tell us that the speculative fiction books overall use more words from the ‘helpful’ lexicon than from the ‘doom’ lexicon.

Again, I briefly want to look at how accurately my method has selected the ‘doom’ instances, seeing if they are actually used in a negative way. In the previous section, there was one book which only contained four sentences with words from the ‘helpful’ lexicon. This does not occur in the corpus when the ‘doom’ lexicon is analyzed, hence I have selected the book with the highest percentage (*The Ascension Factor*) and have selected four sentences that use different words from the ‘doom’ lexicon.

Sentences	Form of ‘doom’
‘This far south, Pandora’s days lasted nearly fourteen hours.’	[‘Pandora’]
‘The search itself might set off the program, put the squeeze on a trigger in his head, a trigger set to destroy them all.’	[‘destroy’]
‘Beyond the scene of this little death in front of him the greater deaths of charred villagers fanned out from the smoking ruins of the Preserve.’	[‘ruins’]
‘Nor will I bring plunder or destruction, not even to avenge life and limb.’	[‘destruction’]

Figure 12: Sentences from *The Ascension Factor* wherein a form of ‘doom’ is applied to AI.

When we look at figure 12, we see that only half of the sentences use the words from the ‘doom’ lexicon in a negative way, namely the second and third sentence. In the second sentence, this becomes clear as it speaks of a ‘trigger’ that will ‘destroy them all’. Next, in the third sentence this is made clear by the use of the words; ‘death’, ‘deaths of charred villagers’, and ‘smoking’, which tell us that the word ‘ruins’ refers to the remnants of a destroyed village. While these two sentence use the words in a negative way, the first and last sentence fail to do this. In the first sentence, the word ‘Pandora’ is the name of the planet and is thus used in a neutral way. The word ‘Pandora’ was included in the ‘doom’ lexicon as it refers to Pandora’s box, which briefly got discussed in chapter one. The tale of Pandora’s box is often used in comparison to AI, because it is believed that when human-like AI is created, it cannot be uncreated, meaning that the consequences cannot be reversed. We should take into account that if this book often mentions the planet’s name, this could be the reason for it having the highest score in figure 11. Lastly, when we look at the last sentence, the word

‘destruction’ is used, but its combination with the word ‘Nor’ strips it of its negativity as it tells the reader that destruction will not happen.

Again, while it is important to point out that the computer has not always selected sentences that use the words from the ‘doom’ scenario in a way that actually implies doom, the computer was still able to provide us with a first impression. It showed us that speculative fiction books use significantly more words from the ‘helpful’ lexicon compared to the ‘doom’ lexicon.

2.3.5 Sentiment Analysis

While the computational analysis of ‘helpful’, ‘doom’ and their other forms did not differentiate between positive and negative use, we can develop a method to search for all of the positive, negative, and neutral words in the corpus. By doing sentiment analysis we can collect data on the positive and negative words and can then make a graph that represents all of the titles in the corpus and their sentiment scores. In this study, the sentiment scores have been calculated using *Vader*, a lexicon and rule-based sentiment analysis tool trained mainly on sentiments expressed in social media.¹¹¹ Next to the computational analysis, I have also performed a close reading to evaluate the accuracy and reliability of the results. Giving the computer the task to collect positive, negative, and neutral words/connotations in the AI sentences provides us with the number shown in the table below. Figure 14 is a visualization of these numbers.

	title	positive	negative	neutral	tokens	positive_perc	negative_perc	neutral_perc
0	Snow_crash	1609	1380	35176	38165	4.215905	3.615878	92.168217
1	The_iron_woman	167	216	5061	5444	3.067597	3.967671	92.964732
2	The_lazarus_effect	1304	1010	23575	25889	5.036888	3.901271	91.061841
3	The_turing_option	1122	706	19745	21573	5.200946	3.272609	91.526445
4	I_Robot	1078	924	16371	18373	5.867305	5.029119	89.103576

Figure 13: Output of the computer’s analysis of positive, negative, and neutral words.

¹¹¹ GitHub, ‘VADER-Sentiment-Analysis’, <https://github.com/cjhutto/vaderSentiment> (August 22, 2023).

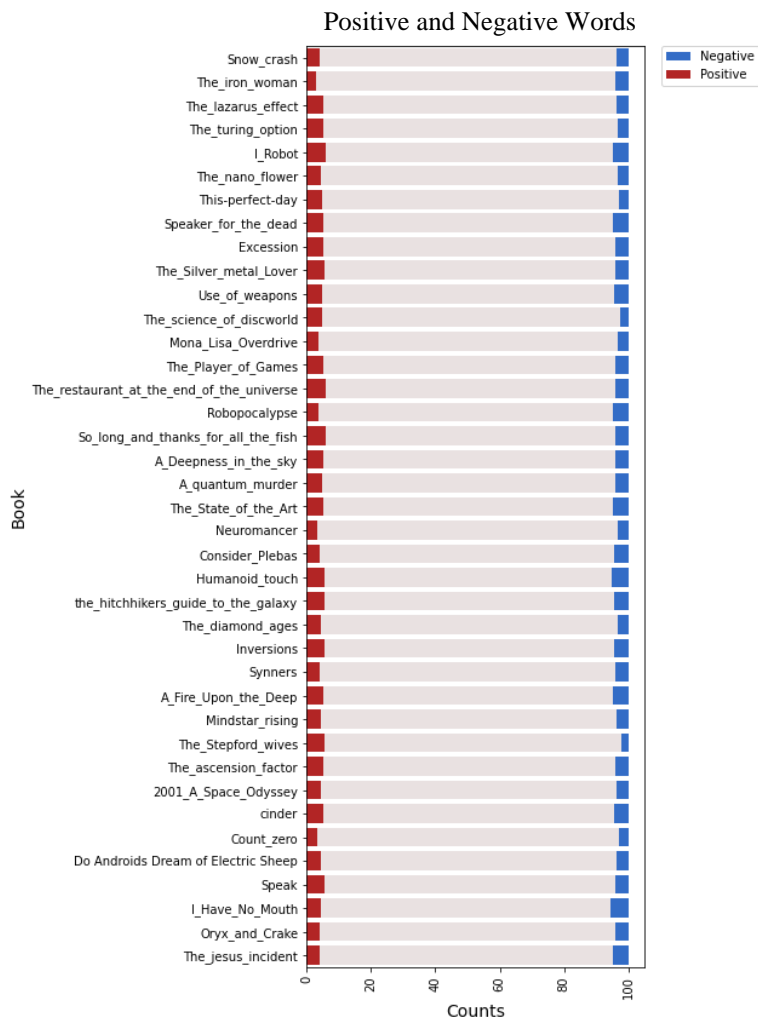


Figure 14: Graph of positive, negative, and neutral words in the AI sentences.

Figure 14 represents all 39 titles and shows percentages of how many positive, negative, and neutral words can be found in each book. The positive words are shown in red, the negative words in blue, and the light red space in between the two colors represents all words with a neutral connotation. From figure 13, we know that *I Robot* exists of around 5.87% positive words, 5.03% negative words, and 89.1% neutral words. These results correspond to the visualization in figure 14, therefore the visualization can be considered accurate. In all the books in the corpus, the words are mostly of a neutral nature. To shed more light on the percentage of the novels, that is positive or negative, I have created a new figure which divides all the books into two categories:

- 1) Books with a higher positive percentage
- 2) Books with a higher negative percentage

The results of this table will be able to tell us if, overall, the books contain more words with a negative connotation or if the positive connotation is predominant.¹¹²

1) Books with a higher positive %	<i>Snow Crash, The Lazarus Effect, The Turing Option, I Robot, The Nano Flower, This Perfect Day, Speaker for the Dead, Excession, The Silver Metal Lover, Use of Weapons, The Science of Discworld, Mona Lisa Overdrive, The Player of Games, The Restaurant at the End of the Universe, So Long and Thanks to All the Fish, A Deepness in the Sky, A Quantum Murder, The State of the Art, Humanoid Touch, The Hitchhikers Guide to the Galaxy, The Diamond Ages, Inversions, Synners, A Fire Upon the Deep, Mindstar Rising, The Stepford Wives, The Ascension Factor, 2001: A Space Odyssey, Cinder, Count Zero, Do Androids Dream of Electric Sheep, Speak, Oryx and Crake, The Jesus Incident</i>
2) Books with a higher negative %	<i>The Iron Woman, Robopocalypse, Neuromancer, Consider Phlebas, I Have No Mouth</i>

Figure 15: Table representing the books with a higher positive or negative percentage in a clear manner.

The table in figure 15 shows us that most books, namely 35 out of 39, use more words with positive connotations in AI sentences. This equals 89.7% of the entire corpus. Further, only 10.3% of the entire corpus uses more words with a negative connotation. This would then mean that most books in the corpus use more words with positive connotations in sentences on AI.

While figure 15 tells us that most books in the corpus mostly use words with positive connotations, the table does not indicate how these words are distributed within the individual texts. Next to overall counts, the development of the plot should be taken into account as a story can often start positively and end in doom, or it can start with a bad event which is resolved in the end. With this in mind, the computer was given the task to create graphs that reflect the entire length of the book linearly, with the beginning of the book on the left and the ending on the right.¹¹³ Figure 16 displays the ‘sentiment arch’ for Jack Williamson’s novel *The Humanoid Touch*.

¹¹² The rest of the output of the computer’s analysis of positive, negative, and neutral words can be found under ‘positive_and_negative_words’ in the repository: GitHub, ‘AI_in_Literature’, https://github.com/peterverhaar/ai_in_literature (August 12, 2023).

¹¹³ This approach is based on: Reagan, A.J., et al., ‘Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs’, *EPJ Data Sci.*, 2017, pp. 1-21.

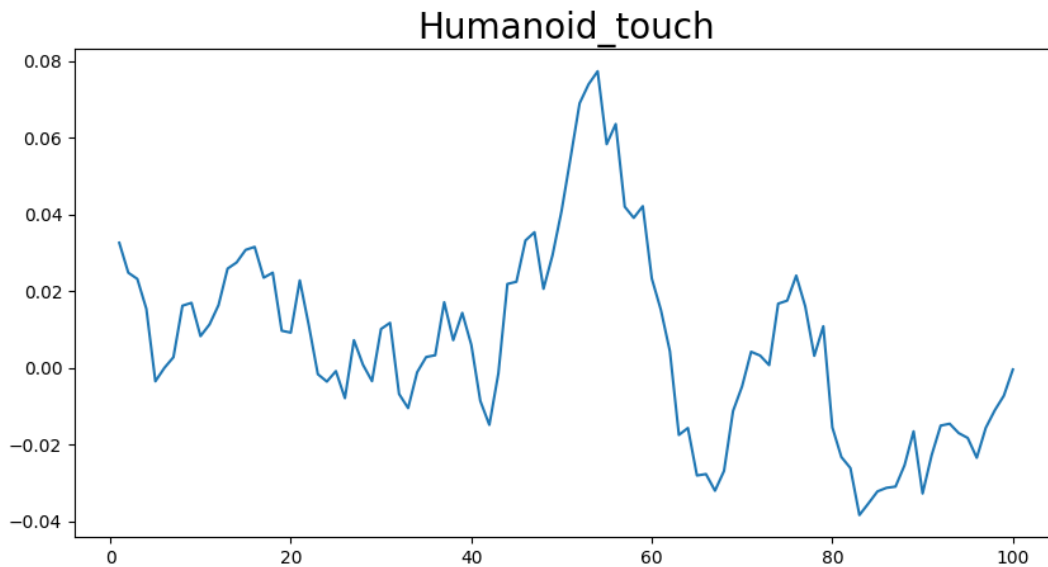


Figure 16: Sentiment analysis of *The Humanoid Touch*

This is a book that, although it progressively gets less positive towards the ending, does stay positive. It starts at a sentiment score of around 0.038 and ends at a score of around 0.005, thus it does portray less positivity in the end, but does not end negatively. We also see that the line does not fluently turn less positive, but that the book instead has many peaks with a larger peak in the middle of the story. Shortly after this peak, we also depict a drop into the negative scores, which tells us that a bad event has happened shortly after a very positive event. This can be proven true by doing a close reading of the drop that should be present at around 67% of the novel. When we go to this percentage in the book, we find the following passage:

‘You won't believe me, shipfolk. The facts will turn you ill. This Keth Kyrone has proved himself inhuman—a merciless monster parading as a man. Only tonight, in the midst of our happy celebration of the humanoids, he forced his way into Vara Vorn.’

‘He found his defenseless victim there, alone and undefended. Fleetmate Chelni Vorn, the young and lovely cousin of Commodore Zoor. She had been with us aboard the *Fortune*, and the whole ship's company had learned to love her. I myself have wished she might have been my own daughter.’

This passage shows us that there has indeed been a negative drop as ‘Keth Kyrone’ is said to have proven himself to be monstrous by killing a young girl that was loved by everyone on the ship.

So, most of the books in the corpus show multiple peaks within the story, but all books have a clear beginning and end. To present the sentiment analysis results of the books in a clear matter, I have chosen to focus on the beginning and the end of every book and assign them to one of the following four categories:

- 1) Books that start and end with a positive score.
- 2) Books that start and end with a negative score.
- 3) Books that start positive and end negative.
- 4) Books that start negative and end positive.

When I had assigned a book to one of these categories, I did include the beginning and end scores in order to show how drastically the scores have changed. Moreover, these differences also offer useful insights into the nature of the books belonging to category one or two. For example, *The Humanoid Touch* belongs to category one, but although it stayed positive, this positive score did decline to a less positive score throughout the book. Assigning every book to one of the four categories gives us the following result:

1) Keep a Positive Score	2) Keep a Negative Score	3) Start Positive, End Negative	4) Start Negative, End Positive
<i>So Long and Thanks for all the Fish</i> (0.015, 0.046)	<i>Robocalypse</i> (-0.017, -0.069)	<i>I Robot</i> (0.023, -0.006)	<i>2001: A Space Odyssey</i> (-0.094, 0.062)
<i>The Turing Option</i> (0.032, 0.068)	<i>Consider Phlebas</i> (-0.02, -0.026)	<i>The Jesus Incident</i> (0.01, -0.03)	<i>Speaker for the Dead</i> (-0.014, 0.046)
<i>The Ascension Factor</i> (0.057, 0.02)	<i>The Iron Woman</i> (-0.034, -0.014)	<i>Use of Weapons</i> (0.014, -0.013)	<i>Speak</i> (-0.002, 0.053)
<i>Neuromancer</i> (0.003, 0.025)	<i>Do Androids Dream of Electric Sheep</i> (-0.052, -0.027)	<i>Count Zero</i> (0.035, -0.007)	<i>The State of the Art</i> (-0.048, 0.038)
<i>The Lazarus Effect</i> (0.039, 0.021)	<i>I Have No Mouth and I Must Scream</i> (-0.036, -0.096)	<i>Snow Crash</i> (0.017, -0.001)	
<i>The Silver Metal Lover</i> (0.05, 0.006)		<i>Inversions</i> (0.06, -0.019)	
<i>Excession</i> (0.026, 0.016)		<i>The Hitchhikers Guide to the Galaxy</i> (0.046, -0.047)	
<i>The Nano Flower</i> (0.046, 0.004)		<i>A Quantum Murder</i> (0.049, -0.014)	
<i>A Deepness in the Sky</i> (0.042, 0.029)		<i>Humanoid Touch</i> (0.033, -0.0004)	
<i>The Restaurant at the end of the Universe</i> (0.021, 0.025)			
<i>Cinder</i> (0.018, 0.003)			
<i>This Perfect Day</i> (0.069, 0.027)			
<i>Mona Lisa Overdrive</i> (0.014, 0.002)			
<i>Mindstar Rising</i> (0.06, 0.009)			
<i>The Science of Discworld</i> (0.082, 0.048)			
<i>A Fire Upon the Deep</i> (0.011, 0.027)			
<i>The Diamond Age</i> (0.069, 0.015)			
<i>Synners</i> (0.016, 0.03)			
<i>The Player of Games</i> (0.089, 0.016)			
<i>The Stepford Wives</i> (0.121, 0.075)			
<i>Oryx and Crake</i> (0.136, 0.107)			

Figure 17: The sentiment analysis scores of the corpus divided into four categories.¹¹⁴

¹¹⁴ All of the graphs that were studied in order to make the following table can be found under ‘Sentiment Analysis.ipynb’: GitHub, ‘AI_in_Literature’, https://github.com/peterverhaar/ai_in_literature (August 20, 2023).

In the figure above, all 39 books from the corpus are divided into categories. The numbers found underneath the book titles in figure 17 represent the score at the start of the book (the number on the left), and the score at the end of the book (the number on the right).¹¹⁵

Moving on to the results, categories one and three were found to be most frequently represented in the corpus, with 21 books in the first category and 9 books in the third. While the large group in the first category tells us that most books on AI begin and end positively, there is more to be said on the changes in positivity scores. When we count the books with a declining positive score, we will note that 15 out of 21 books are less positive in the end compared to the beginning of the book, which equals around 71.4%. So, we can conclude that while most books in the corpus stay positive, around 71.4% of this category does show a decline in the number in which positivity is portrayed in the story.

While we are not certain that these positive instances all refer to AI, as we have noted before that the method applied does not always accurately filters sentences with AI as its subject, all the books in the corpus have AI as its subject or main theme. Hence, we cannot deny that their author's view on AI must have an influence on these scores. Moreover, if linked to AI, categories two and three also portray negativity in the books. These books all show a decline in scores apart from *The Iron Woman* and *Do Androids Dream of Electric Sheep*, which get less negative at the end. The only category that would then fully depict AI positively at the end of the book is category 4, which includes four books. These four books, combined with six books from category 1, and the two books from category 2, form the twelve books that illustrate an increase in positivity at the end of the book. This number equals around 31.8% of the 39 books that were studied. So, most part of the books on AI, namely 68.2%, show a decline in positivity scores, showing us that books with this subject often end either negative or less positive.

As the discussion above only considers the novels' beginning and end, I would also briefly like to consider the middle of the books. Out of the 39 books, the following five books never reach a negative score:

- *The Turing Option* by Harry Harrison and Marvin Minsky
- *The Stepford Wives* by Ira Levin

¹¹⁵ The exact numbers can be found under 'sentiment_scores.csv' in the repository: GitHub, 'AI_in_Literature', https://github.com/peterverhaar/ai_in_literature (August 20, 2023).

- *This Perfect Day* by Ira Levin
- *The Science of Discworld* by Terry Pratchett, Ian Stewart, and Jack Cohen
- *The Diamond Age* by Neal Stephenson

Thus, around 12.8% of the 39 books never presents a significant number of negative words, at least not significant enough to be detected by the computer. If we apply this knowledge to the AI theme of the books, we could argue that only 12.8% of the corpus discusses the subject of AI as something positive throughout the entirety of the book. This would then mean that 87.2% discusses it as a negative phenomenon.

Not being sure if the negativity scores are accurately detected as negative, we will study some of the negative sentences more closely. For this purpose, all of the sentences in which the negative scores were found were brought together in a document. As there is a limit to the space for discussion in this thesis, I have chosen to highlight the two books that keep showing a negative score, so the novels in category two. These books are *Robopocalypse* by Daniel H. Wilson and *Consider Phlebas* by Ian M. Banks. As this section will only study the sentences that scored negatively, these two works enable us to select any sentence. Since the biggest part of the analysis shown above only studied the beginning and end of the novels, we will also focus on these parts in the close analysis of the sentences. From both books, I have tried to avoid cherry picking by selecting the first and the last two sentences of both books in the document, without having prior knowledge on their contents.

Robopocalypse:

First Two Sentences
‘We are a better species for having fought this war.’ (-0.6155)
‘Each walnut-sized robot is lost in the mix as they climb over each other and the whole nightmare jumble of legs and antennae blends together into one seething, murderous mass.’ (-0.7579)
Last Two Sentences
‘The last few years have likely been the only time in human history that we weren’t at war with ourselves.’ (-0.5994)
‘Can it feel guilt or sorrow or shame?’ (-0.8225)

Figure 18: First two and last two sentences with negative scores in *Robopocalypse*

Starting with *Robopocalypse*, we will determine whether the computer has accurately depicted the negativity in these sentences. An analysis of these four sentences points out that only the sentence in the first white area is undoubtedly negative. The sentence in the first

grey area is presumably detected as negative as it talks about fighting and a ‘war’, but it also talks about a ‘better species’ that has evolved from that war, which would be a positive outcome. Similarly, the sentence in the second grey area also discusses war, but also acknowledges that because of this war, humans have not been at war with themselves. This sentence could both be interpreted as positive and negative and should therefore not be put into the negative category as it is more nuanced. Lastly, the sentence in the second white area would be considered negative if it said ‘It cannot feel guilt or sorrow or shame’. Instead, this sentence is shaped as a question and, since it is not answered, I do not see why it should be present in the negative category. As it contains the words ‘guilt’, ‘sorrow’, and ‘shame’, the computer did correctly filter this sentence out as one that contains negative words, but it could not properly consider it in the surrounding context.

Consider Phlebas:

First Two Sentences
‘Accelerating hard, its course a four-dimensional spiral through a blizzard of stars where it knew that only danger waited, it powered into hyperspace on spent engines from an overhauled craft of one class, watched its birthplace disappear astern with battle-damaged sensors from a second, and tested outdated weapon units cannibalised from yet another.’ (-0.7845)
‘The ship did all it could to avoid combat; it kept well away from the routes enemy craft would probably use; it treated every hint of any craft as a confirmed hostile sighting.’ (-0.8225)
Last Two Sentences
‘He reverted shortly afterwards to his earlier role of combat logistics officer and was killed during the twin novae battle for control of arm one-six, towards the end of the war.’ (-0.9246)
‘No sign of wreckage, damage, bodies or any pan of the old changer base was found during the four days that the gcu and its survey teams were permitted to stay.’ (-0.6597)

Figure 19: First two and last two sentences with negative scores in *Consider Phlebas*.

Moving on to *Consider Phlebas*, a study of the negativity of these sentences again raises doubt on the computer’s accuracy. This is the case for the sentences in the white areas. The first sentence does include many words that can be considered negative, namely ‘combat’, ‘enemy’, and ‘hostile’, but it portrays the cautionary stance that the ship takes on to avoid these negative words. As the ship tries to avoid these negative words, we could thus consider its action to be positive. The second sentence also uses negative words, specifically

‘wreckage’, ‘damage’, and ‘bodies’ (as it refers to dead or wounded bodies). However, as it puts ‘No sign of’ in front of these negative words, it turns into a positive sentence as it tells us that these negativities are not present.

So, when we study the computational analysis of negativity in these two novels from category two (‘2: Keep a Negative Score’), we can conclude that while all sentences are made up out of several negative words, they do not always convey a negative message. In *Robocalypse*, 75% of the sentence can be considered as either positive or neither negative nor positive, and in *Consider Phlebas* this can be said about 50% of the sentences. Overall, we can then conclude that the results that were depicted in the four categories are not fully accurate. The computer is thus not yet able to interpret the sentences without errors and still needs a human to analyze them to filter out mistakes. I think that we should not fully write off the results as the computer did succeed in finding negative words in the sentences, it just did not achieve to interpret the sentence its overall message of the sentences. Moreover, as said before, all books in the corpus have AI as its subject or main theme. So, while the results do not accurately show us what part of the negative scores specifically refer to AI as the subject of the sentences, we do know that, regardless of the sentence its subject, all sentences come from books with AI as its subject or theme. This allows us to still make a general claim that the results shown in figure 17 tell us that books with this subject tend to use more negative or less positive words in the end compared to the beginning of the book.

2.3.6 Emotion Mining

The last task that was given to the computer was to study the emotions ‘joy’, ‘fear’, ‘anger’, and ‘surprise’ in the corpus. From these emotions we will specifically study ‘fear’ and ‘joy’ as these emotions can be linked to the opposition which was presented in the first chapter, namely fear of AI and the benefits of AI. The number of sentences that express these emotions can be seen in figure 20 and 21.

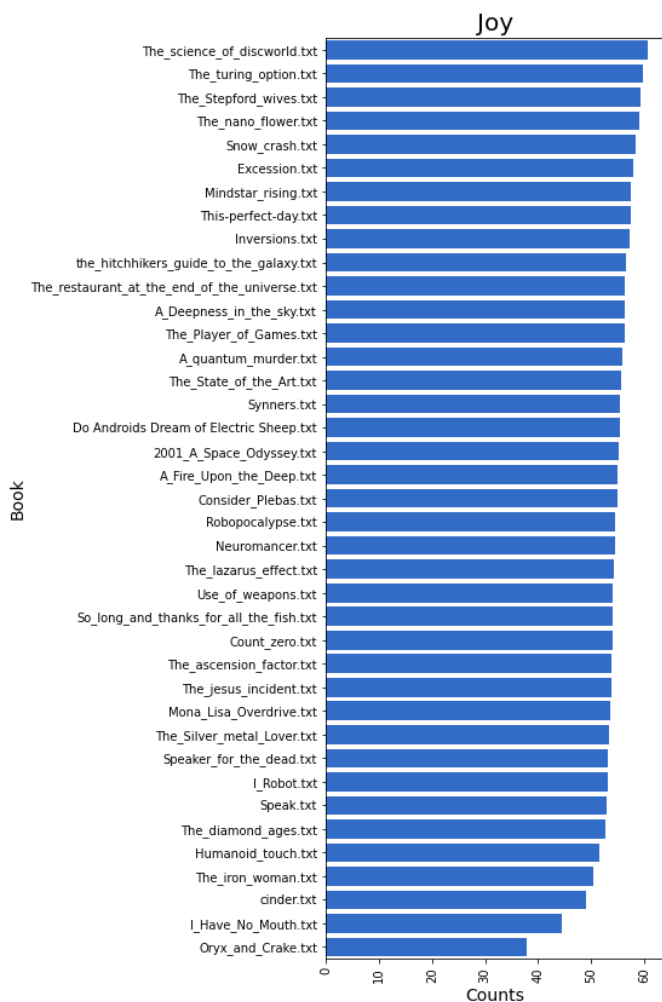


Figure 20: Percentages of sentences that express joy in the corpus.

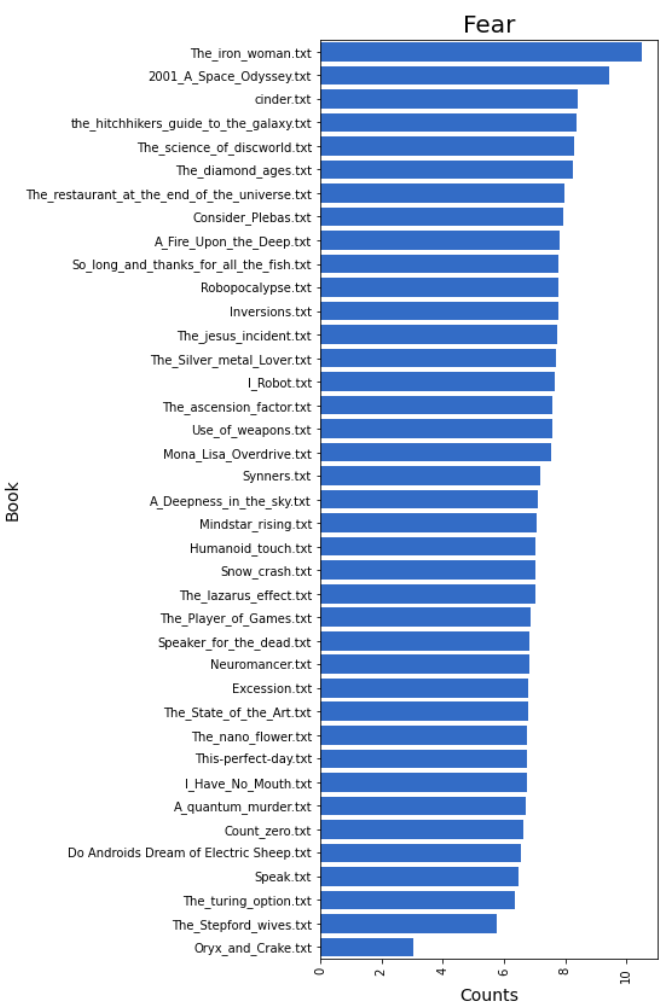


Figure 21: Percentages of sentences that express fear in the corpus.

When we compare the two graphs, we see that in all books, ‘joy’ has the highest counts and is thus the emotion that is most present in all novels. While the counts in figure 20 range between approximately 37% and 57%, the counts in figure 21 range between 3% and 10%, telling us that the emotion ‘joy’ is at least 27% more present in the corpus than ‘fear’. This percentage was calculated by looking at the difference between the highest percentage of ‘fear’ (10%) and the lowest percentage of ‘joy’ (37%).

While comparing figure 20 and 21 allowed us to establish joy as the most prominent emotion in all novels, we are not certain that the computer has correctly interpreted these instances. As we have already noticed a few errors that were made by the computer, we will again closely analyze a few sentences. In order to make figure 20 and 21, the computer has collected sentences which contained joy, and sentences which retained fear. These sentences will help us decide how accurately the computer has interpreted both emotions in the novels. As the length of this thesis has to be taken into account, we will again focus on two novels, which are *The Science of Discworld* and *The Iron Woman*. This choice is somewhat random as the books were chosen due to their placement at the top of figure 11 and 12, but I did not

have any prior knowledge as to what results they might provide. Also, similarly to before, I have chosen to study the first and last sentences for consistency. For both books, I have selected one joy sentence and fear sentence from the beginning of the book, and one joy sentence and fear sentence from its ending. In total, this gives us four sentences to study per book.

The Science of Discworld:

First Two Sentences
‘The story starts here ... once upon a time, there was Discworld.’ (Joy)
‘Discworld does not run on scientific lines.’ (Fear)
Last Two Sentences
‘As above, so below.’ (Joy)
‘The Librarian was hiding behind a chair.’ (Fear)

Figure 22: The first sentence and last sentence that express joy and fear in *The Science of Discworld*

Starting with *The Science of Discworld*, we can already see a problem when we study the two sentences on joy in the grey areas. While they overall seem to be positive sentences, they do not express a form of joy when they are interpreted by a human. Opposite to the joy sentences, the fear sentences in the white areas do contain one sentence that expresses a form of fear. While the first sentence does not express a form of fear, the second one does by describing a librarian who hides behind a chair. The emotion of fear is not clearly named, but when a human interprets this sentence, we would often consider the act of ‘hiding behind something’ as a result of fear. However, when we read through the context surrounding this sentence, we will notice that this sentence also does not express fear:

‘‘Okay, you can come out,’ he said. ‘They’ve lost interest.’

The Librarian was hiding behind a chair. The orangutan took university discipline seriously, even though he was capable of clapping someone on both ears and forcing his brain down his nose.’

Through this excerpt, we come to understand that the librarian is an orangutan that acknowledges that it is capable of self-defense and is only hiding as it is a university rule. This tells us that the orangutan is never actually threatened and is thus not hiding because of fear, but instead because it follows discipline. So, all four of the sentences in figure 22 fail to

express the emotions that the computer searched for in this book, which tells us that the computer is not able to accurately interpret emotions.

The Iron Woman:

First Two Sentences
‘Lucy was walking home, between the reed banks, along the marsh road, when it started to happen.’ (Joy)
‘School was over and the Easter holidays had begun.’ (Fear)
Last Two Sentences
‘It was not a skylark.’ (Joy)
‘They both looked up into the blue, gazing and listening.’ (Fear)

Figure 23: The first sentence and last sentence that express joy and fear in *The Iron Woman*

To check if the computer really is unable to detect emotions or if we have coincidentally chosen four sentences that were inaccurately analyzed, we will study one more book, namely *The Iron Woman*. Figure 23 shows us that the first two sentences already contain an error. In the book, the fear sentence is followed up by the joy sentences, so:

‘School was over and the Easter holidays had begun. Lucy was walking home, between the reed banks, along the marsh road, when it started to happen.’

While the computer has interpreted school being over and the start of the holidays as a fearful event, most people would consider this to be a joyful event. Further, in the sentence in which the computer did detect joy, the word ‘it’ instead refers to a frightening event. In chapter one, we read how, when Lucy walks home, she always sees the animals behave strangely in the polluted water. She observes how animals suffer from painful deaths, which sounds fearful or at least it does not bring forth feelings of joy. The computer has thus assigned these sentences an opposite emotion as to what they actually convey.

Further, the last two sentences can be found on the very last page of the book where the protagonist Lucy and three others are noticing a sound becoming stronger, stating that:

‘It was not the faint sound of the creatures crying. It was music of a kind, from far off, far up. They both looked up into the blue, gazing and listening. It was not a skylark.’

Considering its context, the fear sentence seems to be describing a more joyful situation as the characters are ‘gazing’ into a ‘blue’ sky and ‘listening’ to some sort of ‘music’. As the

whole plot of the novel sets out to solve the pollution that is portrayed in the first chapter with the suffering of the animals, this sentence actually points out to the reader that the goal is achieved. The word ‘blue’ is associated with a clear sky, which indicated joy and a lack of pollution. Moreover, the word ‘music’ is used as an opposite of ‘creatures crying’, which tells us that the animals are no longer suffering due to pollution.

Lastly, we have the joy sentence ‘It was not a skylark’, which does not clearly express the emotion joy, but instead seems neutral, and can therefore be seen as an inaccuracy of the computer. However, if one has read the entire novel, this sentence can be considered a positive, or even joyful, one. The skylark appears multiple times throughout the novel and sings when it presents itself. It appeared in times of trouble when the pollution was still an issue that the protagonist was fighting to solve. The last sentence saying ‘It was not a skylark’ can thus tell us that the skylark was a symbol for the pollution and has now stopped appearing as the pollution issue is resolved, which then implies that the last sentence can at least be considered a positive one.

All in all, the eight sentences that were studied mostly can be linked to an emotion when their context is considered, but they do not clearly express an emotion when they are studied separately. We can conclude that the computer has failed to accurately analyze joy and fear in the sentences, as there are only two sentences that we can consider to be interpreted correctly, but even this is questionable. These instances refer to 1) the librarian hiding behind a chair, which does show a sign of fear but this can be ruled out when the context is considered; and 2) the last sentence on the skylark, which could express joy when it is read as a symbol. However, as the computer was not asked to look for symbols in the text, I find it hard to believe that it interpreted what this symbol stood for. Computers are thus not yet able to accurately detect emotions in texts, therefore the results shown in ‘Emotion Mining’ cannot be taken into account when we try to answer the research question.

2.4 Discussion of the results

If we look at all the analyses collectively, we can argue that they all delve into the challenges and constraints of computer-driven analysis of the 39 books on AI in the corpus. Through the study of the AI sentences, animacy, ‘helpful’ lexicon, ‘doom’ lexicon, sentiment analysis, and emotion mining, the computer’s capabilities and limitations have become apparent, and the ongoing need for human intervention and interpretation has been highlighted. However,

despite the limitations, the computational analysis still allowed us to make general claims on AI.

The first study on AI sentences detected that errors emerge when the computer identifies AI in sentences in the corpus. These errors arise due to misinterpretation of context and content. Despite the inaccuracies, the section still affirmed the study's value as it focuses on books with AI as its subjects, thus the results still provide insights into AI-focused fiction.

Next, the section on animacy focused on the portrayal of human-like motion and emotion in various books from the corpus. When the computer was asked to detect animated words in the books, it only succeeded in looking for animacy in two of the books from the corpus. This was solved by manually selecting sentences on AI. By studying the various sentences, we notice that many sentences depict AI in a human-like way or showcased its supportive role. These findings imply that speculative fiction frequently depicts AI as having human-like qualities.

The third study on the 'helpful' lexicon has shown us that the task of discerning AI as a 'helpful' being, is still a difficult for the computer. It especially has difficulty with discerning whether the word from the 'helpful' lexicon is used in a positive way. The lack of differentiation between positive and negative words complicates achieving a conclusion on AI portrayal of a helpful thing or being. However, when we see it as a first impression, we can determine that these words are often used in speculative fiction books that have AI as its subject.

Additionally, a study of the 'doom' lexicon allows us to compare its results to the 'helpful' lexicon. While the computer also has shown difficulty with interpreting the words from the 'doom' lexicon as negative, the section does tell us that words from the 'helpful' lexicon are more frequently used throughout the corpus. This suggests that speculative fiction tends to focus more on positivity in AI-focused-books than doomsday scenarios.

Moving on to sentiment analysis, it again highlighted the computer's limitations in accurately interpreting negativity in sentences. Although the computer identified negative words, it failed to grasp the overall message of the sentences. Despite this, it is acknowledged that the presence of negative words in AI-themed books implies a broader pattern, but the specific negative message of the sentences remains elusive without human interpretation.

Further, the emotion mining section showed that the computer struggles to detect and understand emotions in the sentences that were discussed. While some instances may convey emotions when we consider its context, the computer's analysis failed to accurately identify the emotions 'joy' and 'fear' in the context, but instead focused on the isolated words.

Consequently, the computer's emotional analysis remains unreliable, which casts doubt on if we could regard the results presented as useful.

So, when we go back to the research question, which considers the computer's ability to answer the question: How is AI represented in speculative fiction published between 1950 and 2015? We can conclude that the computer showed difficulty to provide the results to answer this question in all of the studied sections. What we have been able to determine is that speculative fiction often represents AI as things or beings with human like motion and emotion. Additionally, the genre tends to use more words that describe a form of helpful rather than a form of doom. This positivity is emphasized by the genre mostly portraying books with a positive ending, although the positivity scores have often shown to decline at the end of the books. This first impression thus implies that the speculative fiction genre focuses more on positivity when discussing AI.

Chapter 3: Conclusion

The goal of this thesis was to study how AI is represented in speculative fiction published between 1950 and 2015, and to see if computational capabilities are able to help us answer this question. The first chapter set out to introduce AI its definitions, history, capabilities, complications, and the views that people hold towards AI. By depicting AI in its many forms and with its many qualities, we looked at AI in a nuanced way, and came to understand what portrayals of AI we could run into when we would delve into the case study presented in chapter two. Chapter one also showed us how far technology has developed nowadays. The capabilities of technology were further shown in the case study as it revealed to us how helpful computational methods were in providing us with the data needed to answer the research question.

The results of the case study were divided into six different sections:

- 1) A study of the AI sentences.
- 2) A study of animacy.
- 3) A study of the word ‘helpful’ and its synonyms in relation to AI.
- 4) A study of the word ‘doom’ and its synonyms in relation to AI.
- 5) A study of the positive and negative words in the books and their development.
- 6) A study of the emotions present in the books, focusing on ‘Joy’ and ‘Fear’.

When we collectively studied these results, we could contend that they all focus on the difficulties and limitations that computer-driven analysis still shows. While chapter one has given us an insight into technology its development, the case study has shown that computational analysis has difficulty interpreting the context surrounding a detected word. Nonetheless, it was able to point us in a direction, permitting us to make generalized claims on speculative fiction.

With the help of computational analysis, we have discovered that the language used in this genre frequently uses words that describe a form of ‘helpful’ as opposed to forms of ‘doom’. Although the positivity scores have frequently demonstrated to fall at the end of the books, the genre emphasizes positivity by depicting books that generally have a happy ending. This provides us with the initial impression that speculative fiction tends to emphasize on positivity more. The research also included a section on joy, but as the computer failed to interpret these emotions correctly, we cannot use the results of that section to make claims on the genre. While the computer also showed difficulty interpreting the

context of the words from the ‘helpful’ and ‘doom’ lexicon, it was able to detect all of the words from these lexicons in the texts. So, while the shortcomings in computational analysis did not allow us to make discoveries relating to AI specifically as helpful or terrible beings, it did provide insight on the genre that represents AI. When we look back at chapter one, which portrayed the clear opposition between AI’s benefits and it possibly being our downfall, we can conclude that speculative fiction overall focuses more on a positive message and a happy ending.

Furthermore, in chapter one, the notion was introduced that our fear of AI might originate from fiction, including books and movies. Rogers believed that a pre-existing dread for AI has resulted in a negative portrayal of AI in fiction. The significance of this narrative is emphasized by Cave and Dihal, who discussed how narratives shape the goals of AI developers, public acceptance, and policy making. Their research also explained that people tend to recognize both the positive and negative notions about AI. Although the results presented in chapter two made assertions on speculative fiction in general, rather than on AI specifically, we are now aware that these AI-themed books tend to utilize more positive words. This contrasts with Rogers’ perspective, which suggests that literature conveys a more negative message. The results also do not completely match Cave and Dihal’s findings. While we were able to find both positive and negative scores in the corpus, the positive scores outweighed the negative scores. This implies that speculative fiction mostly has a positive effect on the goals of AI developers, public acceptance, and policy making. While books can still play a role in our fear of AI, it does not appear to be a significant factor in stoking this fear.

Moreover, chapter one also mentioned prior studies by the *House of lords* committee and *Hermann* that have already made claims about ‘animacy’ and ‘humanness’, without using these specific terms. Chapter 2.3.2 has proven that these claims are true. While the *House of Lords* and *Hermann* studied this aspect in fiction and science-fiction, this study focused on speculative fiction in particular. The *House of Lords* committee stated that AI is often portrayed as humanoid in fiction and this research has shown that AI is indeed often portrayed as a thing or being that can move and talk in a human-like manner. In the sentences that were studied, human-like movement occurred more frequently than human-like speech. Further, Hermann dived deeper into the idea of humanoid AI by studying if fiction portrayed them as things that exceed the intended capabilities of their creation, such as feeling emotion. The corpus has also shown to include the representation of AI that experience emotion through a close reading of *The Player of Games*.

While we are now aware that depicting AI in a human-like way is often done in speculative fiction, we do not yet know if this portrayal is linked to the positivity that is mostly present in speculative fiction works. We can wonder if the books dominant use of positive words and words from the ‘helpful’ lexicon are mostly used when books include human-like AI. This would be an interesting case study for future research. Moreover, as technology rapidly develops, future research will presumably be able to make use of better computational analysis methods and will then be able to use this research as a starting point. If computers will acquire better interpretation skills, future research will be able to accurately study the words from the ‘helpful’ and ‘doom’ lexicon, and positive and negative scores in relation to AI.

Even though this research has its flaws due to the computers’ lack of interpretation skills, it is still a useful contribution to existing research on AI in fiction. It has combined methods and ideas of existing research to study a specific time period (1950-2015) in the speculative fiction genre. While some complications remain, the study accomplished to attain new insights into the speculative fiction genre and the capabilities of computational analysis. We now know that the genre tends to convey a more positive message. As the research shows errors in the computers’ interpretation, which cannot be solved yet, this research provides a starting point for future research.

Bibliography

- Ardanuy, M.C., et al., 'Living Machines: A study of atypical animacy', *Proceedings of the 28th International Conference of Computational Linguistics*, 2020, pp. 4534-4545.
- Åsberg, C., 'Feminist posthumanities', *Posthuman Glossary*, 1 (2018), pp. 157-160.
- Atwood, M., 'Dire Cartographies: The Roads to Ustopia', *Other Worlds: SF and the Human Imagination* (New York: Doubleday, 2011), pp. 89-127.
- Beentjes, M., '19th Century Issues in the Modern Age: The Posthuman, Transhuman and Female Posthuman in Jeanette Winterson's Reimagination of Frankenstein' (University of Amsterdam, 2022), pp 1-33.
- Borry, P & G. Matthijs, 'Designer Babies', *The Human Recipe: Understanding Your Genes in Today's Society* (Leuven: Leuven University Press, 2016), pp. 93-102.
- Braidotti, R., 'Posthuman critical theory', *Posthuman Glossary*, 1 (2018), pp. 339-340.
- Braidotti, R., *The Posthuman* (Cambridge: Polity, 2013).
- Bringsjord, S., et al., 'Creativity, the Turing Test, and the (Better) Lovelace Test', *Minds and Machines*, 11 (2001), pp. 3-27.
- Butler, S., *Erewhon*, (New York: Octagon Books, 1968).
- Cave, S. & Dihal, K., 'Hopes and fears for intelligent machines in fiction and reality', 1 (2019), pp. 74–78.
- CEPR, 'AI and the paperclip problem', <https://cepr.org/voxeu/columns/ai-and-paperclip-problem> (June 28, 2023).
- Ceruzzi, P. E., 'Moore's Law and Technological Determinism: Reflections on the History of Technology', *Technology and Culture*, 46 (2005), pp. 584-593.
- Darwin, C., *On the Origin of Species: by Means of Natural Selection, or Preservation of Favored Races in the Struggle for Life* (London: John Murray, 1859).
- Garland, A., *Ex machina* (Santa Monica, CA: Lionsgate, 2015).
- GitHub, 'AI_in_Literature', https://github.com/peterverhaar/ai_in_literature (August 12, 2023).
- GitHub, 'VADER-Sentiment-Analysis', <https://github.com/cjhutto/vaderSentiment> (August 22, 2023).
- Good, J.I., 'Speculations Concerning the First Intelligent Machine', *Advances in Computers*, 6 (1965), pp. 31-88.
- Hanson Robotics, 'Sophia', <https://www.hansonrobotics.com/sophia/> (May 5, 2023).

Hanson Robotics Limited, 'Sophia the Robot by Hanson Robotics', *Youtube*, 2018, <https://www.youtube.com/watch?v=BhU9hOo5Cuc> (May 5, 2023).

Harari, Y., *Homo Deus: a brief history of tomorrow* (New York: HarperCollins, 2017).

Hayles, K.N., *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatic*, (Chicago: The University of Chicago Press, 1999).

Hermann, I., 'Künstliche Intelligenz in der science-fiction: zwischen magie und technik', *FifF-Kommun*, 4 (2020), pp. 12–17.

Hermans, S., *(R)Evolutie* (Nederland: VPRO Cinema, 2021).

House of Lords, *AI in the UK: ready, willing and able?*, House of Lords: Select Committee on Artificial Intelligence (Scientific Research: An Academic Publisher, 2018), <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.

iRobot, 'Roomba@ Robot vacuums', https://www.irobot.com/en_US/roomba.html (June 4, 2023).

Kurzweil, R., *The Singularity is near: When Humans Transcend Biology* (New York: Viking, 2005).

Liu, P., et al., 'Application of Artificial Intelligence in Medicine: An Overview', *Current Medical Science*, 41 (2021), pp. 1105-1115.

Lucasfilm Ltd.; 20th Century Fox, *Star Wars original trilogy* (San Francisco: Lucasfilm, Twentieth Century Fox Home Entertainment, 2013).

Make Use of (MUD), '5 Ways AI Can Help Fight Climate Change', <https://www.makeuseof.com/ways-ai-help-fight-climate-change/> (June 6, 2023).

Manning, C., 'Artificial Intelligence Definitions', *HAI* (Stanford University, 2020).

Marovich, L. A., "'Let Her Have Brains Too': Commercial Networks, Public Relations, and the Business of Invention", *Business and Economic History*, 27 (1998), pp. 155-157.

Negroponce, N., *Being Digital*, (London: Hodder & Stoughton, 1995).

OpenAI, 'Introducing ChatGPT', <https://openai.com/blog/chatgpt> (May 28, 2023).

Princeton University, 'WordNet', <https://wordnet.princeton.edu/> (August 12, 2023).

Reagan, A.J., et al., 'Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs', *EPJ Data Sci.*, 2017, pp. 1-21.

Rogers, G., *Speculation: A cultural History from Aristotle to AI*, (New York: Columbia University Press, 2021).

Scott, J., 'Masculinities and sex workers', *Routledge international handbook of masculinity studies* (London: Routledge, 2020), pp 271-280.

- Self, J., 'Darwin Among the Machines', *Architectural Design*, 83 (2013), pp. 66-71.
- Smith, M.R. & L. Marx, *Does Technology Drive History?: The Dilemma of Technological Determinism* (Cambridge: The MIT Press, 1994).
- Stokel-Walker, C. & R. Van Noorden, 'What ChatGPT and generative AI mean for science', *Nature*, 614 (2023) pp. 214-216.
- Szollosy, M., 'Freud, Frankenstein and Our Fear of Robots: Projection in Our Cultural Perception of Technology', *AI & Society*, 32 (2017), pp. 433-439.
- Thunström, O & S.Steingrimsson, 'Can GPT-3 write an academic paper on itself, with minimal human input?', *HAL Open Science* (2022), preprint version, pp. 1-7.
- Turing, A.M., 'Computing Machinery and Intelligence', *Mind*, 59 (1950), pp. 433-460.
- Universiteit Leiden, 'Computational Drug Discovery: Key publications', <https://www.universiteitleiden.nl/en/science/drug-research/drug-discovery-and-safety/computational-drug-discovery/key-publications> (July 8, 2023).
- Westen, G.J.P. van, et al., 'Proteochemometric modeling as a tool to design selective compounds and for extrapolating the novel targets', *MedChemComm*, 2 (2011), pp. 16-30.
- Winner, L., 'Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology', *Science, Technology & Human Values*, 18 (1993), pp. 362-378.
- Winnow, 'The business case for using food waste technology', <https://info.winnowsolutions.com/food-waste-technology> (June 30, 2023).
- Winterson, J., 'Jeanette Winterson and Victoria Turk: 12 Bytes', Interview by Victoria Turk, *YouTube*, 2021, <https://www.youtube.com/watch?v=MuhU82NIdyE&t=319s> (June 5, 2023).
- Winterson, J., *12 Bytes* (London: Jonathan Cape, 2021).
- Zarkadakis, G., 'From Bletchley Park to Google Campus', *In Our Own Image* (Cambridge: Pegasus, 2015), pp. 211-229.
- Zarkadakis, G., 'Machines that Think', *In Our Own Image* (Cambridge: Pegasus, 2015), pp. 230-250.

Appendices

Appendix 1

The corpus (1950-2015)

1. Adams, D. *The Hitchhiker's Guide to the Galaxy*.
2. Adams, D. *The restaurant at the end of the universe*.
3. Adams, D. *So long, and thanks for all the fish*.
4. Asimov, Isaac. *I, Robot*.
5. Atwood, M. *Oryx and Crake*.
6. Clarke, A. C. *2001: a space odyssey*.
7. Banks, I. M. *Consider Phlebas*.
8. Banks, I. M. *The Player of Games*.
9. Banks, I. M. *Use of Weapons*.
10. Banks, I. M. *The State of the Art*.
11. Banks, I. M. *Excession*.
12. Banks, I. M. *Inversions*.
13. Cadigan, P. *Synners*.
14. Card, O. S. *Speaker for the dead*.
15. Dick, P. K. *Do androids dream of electric sheep?*
16. Ellison, H. *I Have No Mouth and I Must Scream*.
17. Gibson, W. *Neuromancer*.
18. Gibson, W. *Count Zero*.
19. Gibson, W. *Mona Lisa overdrive*.
20. Hamilton, P. F. *Mindstar Rising*.
21. Hamilton, P. F. *A Quantum Murder*.
22. Hamilton, P. F. *The Nano Flower*.
23. Hall, L. *Speak*.
24. Harrison, H. & Minsky, M. *The Turing option*.
25. Herbert, F. & Ransom, B. *The Jesus incident*.
26. Herbert, F. & Ransom, B. *The Lazarus effect*.
27. Herbert, F. & Ransom, B. *The ascension factor*.
28. Hughes, T. *The Iron Woman*.
29. Lee, T. *The silver metal lover*.
30. Levin, I. *This Perfect Day*.
31. Levin, I. *The Stepford Wives*.
32. Marissa, M. *Cinder*.
33. Pratchett, T., Stewart, I. & Cohen, J. *The science of Discworld*.
34. Stephenson, N. *Snow crash*.
35. Stephenson, N. *The diamond age*.
36. Vinge, V. *A Fire upon the Deep*.
37. Vinge, V. *A Deepness in the Sky*.
38. Williamson, J. *The humanoid touch*.
39. Wilson, D. *Robopocalypse*.

Appendix 2

The 'helpful' lexicon:

benefit	serve
benefited	served
benefits	serving
benefitting	backing
expanding	confidence
beneficial	facilitate
service	facilitates
help	facilitated
helper	facilitating
helping	invaluable
helps	helpful
helped	helpfully
profit	assistant
profits	precious
profited	processor
worthful	useful
wonderfully	usefully
utilitarian	welfare
useful	well-being
avail	incredibly
priceless	support
assistance	supports
assistant	supported
valuable	supporting
assist	aid
assisting	aide
assists	assist
assisted	assisted
improve	assists
improves	assisting
improved	function
improving	gain
perfectly	gains
utile	gained
serves	gaining

Appendix 3

The 'doom' lexicon:

pandora	carried off
kill off	carrying off
extirpate	ruin
extirpates	ruined
extirpated	ruins
extirpating	ruining
demolish	eradicate
demolishes	eradicated
demolishing	eradicates
day of reckoning	eradicating
wipe out	eradications
wipes out	decimate
wiped out	extinction
wiping out	disintegration
root out	annihilate
extinguish	devastation
extinguished	uproot
extinguishes	demolition
extinguishing	exterminate
end of the world	doom
doomsday	destroy
extermination	destroys
exterminate	destroyed
exterminates	destroying
exterminated	eliminate
exterminating	eliminates
obliteration	eliminated
obliterate	eliminating
obliterates	annihilation
obliterated	annihilate
obliterating	annihilates
destruct	annihilated
destructs	annihilating
destructed	condemn
destructing	destruction
carry off	wipeout
carries off	