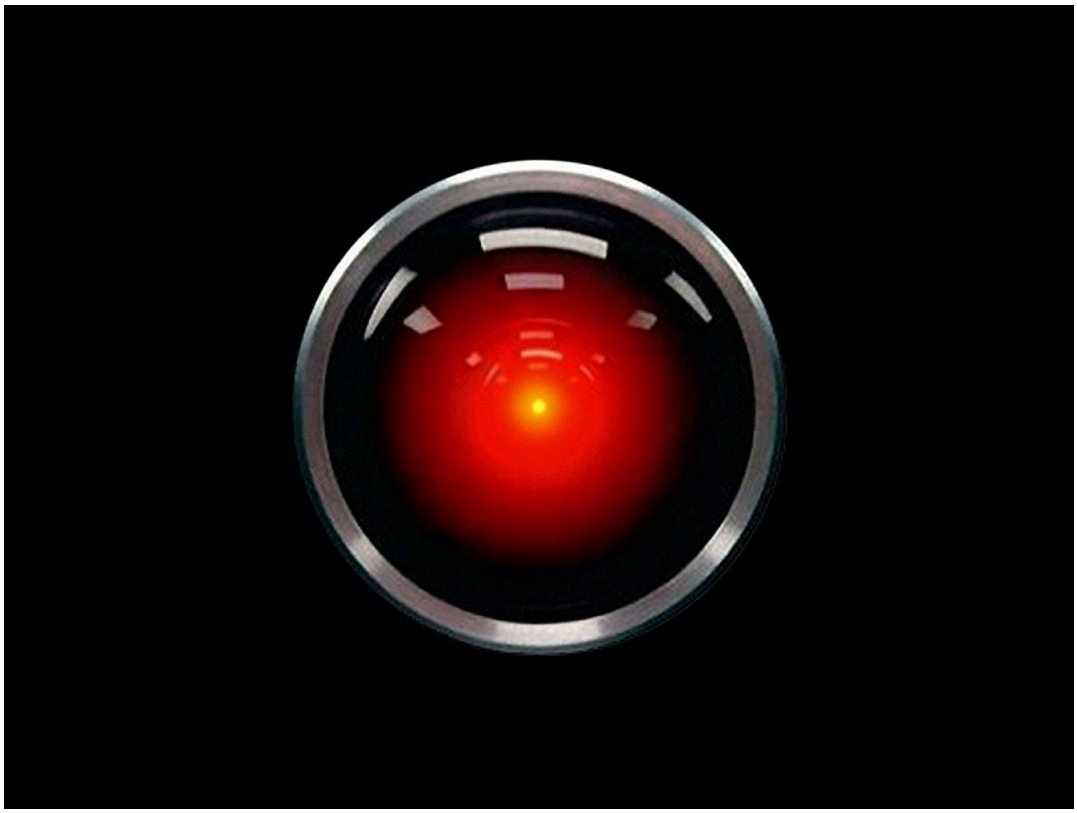
BETWEEN HYPE AND SCAREMONGERING IN AN AGE OF AI



A philosophical inquiry into three scenarios on the current and future state of AI.

15-09-2020

Master's thesis for MSc. in Business Administration and Philosophy

Characters / pages: 147.333 / 65

Supervised by Ole Bjerg

Written by Bjartur Dal-Christiansen (93948)

Abstract

Artificial intelligence is becoming increasingly important in our times in both business and society. It is well established that artificial intelligence will continue to have large-scale consequences across many levels of society both in the near and far future.

This study uncovers the story of technological progress within the narrative of artificial intelligence, as it appears in selected present and future scenarios. The research objective seeks to map out three different scenarios regarding the current and future state of AI. In this context, Pragmatic AI is defined as the business use case for AI technology, Death by Banality is defined as an overly pessimistic attitude to the future of AI, and Digital Utopia is defined as an overly optimistic attitude to the future potential of AI.

To discuss and analyze the respective scenarios I apply Charles Eisenstein's overarching philosophical concept of the Story of Separation and the related concepts of technology, science, reason, and control. The Story of Separation both challenges and critiques established narratives and assumptions about artificial intelligence. In conclusion, I show that rooted deep within the story of artificial intelligence in both the current Pragmatic as well as both future scenarios, there is a story about technological progress based on an even deeper Story of Separation.

Table of Contents

1. Introduction

Whether we like it or not, artificial intelligence (AI) is going to have broad-scale implications across all levels of society (Brynjolfsson & McAfee, 2015; Davenport, 2018). It will change the world as we know it. Whether it be in business, journalism, healthcare, or whatever sector of society one chooses to focus on, AI is going to change it for good. With these impacts, it can sometimes be difficult to keep track of what is actually happening at the ground floors of the AI trenches, what is cutting edge research not yet implemented on a broader scale, and what is science fiction or just merely wishful thinking.

The point of departure for this thesis began with the simple question: What is artificial intelligence? While the question is simple, the answer is not uncomplicated. A simple introductory philosophical definition of AI must deal with what is intelligent and what is not, and what is artificial as opposed to natural. Also, one must look at how AI actually functions in practice as well as what researchers envision AI becoming in the future. These distinctions in themselves are not a simple matter, nor are these easy concepts to clarify, although that is not what this thesis seeks out to do. What I ask instead is this: What are the visions on the future (and current state) of AI technology? Where is it taking us and how are we going to get there? If the future of "human-level" artificial intelligence corresponds purely to the calculative, rational, logical intelligence of artificial intelligence, is that future something we want to keep working towards?

In the last couple of decades, AI has received increasing attention in both print and electronic media, books, scientific literature, as well as in the general public's consciousness. One of the reasons for the recent developments in AI is that computing power and storage capacity has increased dramatically the last many years and that the amount of data collectively being produced all around the Earth has grown immensely over the years (Marr & Ward, 2019, p. 5). One source claims that by 2020, "there will be 40 [times] more bytes of data than there are stars in the observable universe" (Domo, 2019).[1] The developments in computing power, storage

[1] The concrete number of bytes in existence remains uncertain yet was included for illustrative purposes only. For instance, a World Economic Forum article from 2019 referred to a Raconteur infographic, which in turn referred to a PwC report from 2018 which itself referred to a Computer Weekly article from 2012 which finally landed me at the source of an International Data Corporation (IDC) "Digital Universe study" from 2010 that predicted 2020 levels of data in the known "data universe". This number even seems to be wrongly quoted in all subsequent mentions as 40 Zettabytes (one Zettabyte is 1.000.000.000.000 or 10^{12} Gigabytes) instead of 35 ZB as in the original. Bughin et al. (2018) quote another IDC from 2017 predicting 163 Zettabytes or (1 trillion gigabytes) to be created in 2025 (p. 5).
Suffice it to say, though, that the amount of data being produced collectively around the world at any given moment is astronomical in the most literal sense.

capacity, along with the increased data production and collection has made possible artificial intelligence systems capable of solving complex analyses of the large amounts of data that an intelligent AI needs in order to solve different tasks, analyze large amounts of data, make predictions, or help partially or fully automate decision-making (Burton, Stein & Jensen, 2019; Kulager, 2019).

Peer-reviewed journal articles, fiction and non-fiction books, as well as articles in print and electronic media on the topic of AI are as diverse and multifaceted as one can imagine. While ranging from the positive effects of AI in healthcare (Vincent, 2020b); to how Microsoft created an allegedly racist Twitter chatbot (Hunt, 2016); to your smartwatch tracking the quality of your sleep (Søndergaard, 2020); to an Netflix' suggestion algorithm helping you decide what to watch (Marr and Ward, 2019, pp. 161-166); and almost anything imaginable involving data and analytics (Davenport, 2018), there are some noticeable trends or currents when dealing with the topic of AI.[2] One of the main topics of the AI discussion is that of artificial general intelligence (AGI). In popular media it is often presented that if we do not take enormous care when designing an AGI, this 'true AI' might just become so powerful that it deems us, human beings, as surplus to requirements (see, e.g., Holley, 2015; McFarland, 2014; Cellan-Jones, 2014), and while one could devote an entire master's thesis on the reasons for this particular angle being so popular (i.e., does it generate more clicks than other, less "sexy" [Davenport, 2019, p. 7] angles?), this approach to the problem is not the focus of this thesis.

Davenport (2018) defines AI as computers solving narrowly defined tasks making use of capacities such as "knowledge, insight, and perception" previously thought to only belong to the realm of humans (p. 9), while Marr and Ward (2018) define it as "the ability of computer systems or machines to display intelligent behavior that allows them to act and learn autonomously" (p. 3). This capability concludes in applying an algorithm, (i.e., a set of calculation rules) to data and through that process making decisions or predicting an outcome (Marr & Ward, 2019, p. 4). This is what is commonly referred to as 'narrow' AI and for the purposes of establishing a working definition shall be taken to refer to specific task- or problem-solving computer-based algorithms that help humans make better data-driven decisions based on facts.

Both narrow and general AI, are at the moment still critically dependent on good quality data from which to generate their desired output. Therefore, one of the most crucial discussions around the ethics of AI is that of 'data bias' (see, e.g. Hao, 2019; Manyika, 2019). An example of

[2] I present these in further detail throughout this introduction as well as in section 2.3 in Chapter 2.

this is one of, if not the, most powerful artificial intelligence of its kind ever created, a text-generating AI called GPT-3 created by the firm OpenAI. It is said to have read virtually all written content on the Internet and is said to be able to write just as well or even better than a human being (Vincent, 2020a). The problem is that GPT-3 reflects the implicit bias in all of the written content that is already in existence on the Internet. If much of that is prejudiced towards Muslims as terrorists, then this is the angle that GPT-3 will choose for its writing (Kulager, 2020). A more in-depth discussion of data bias shall be presented in Chapter 3.1.2.

For the context of this thesis, it is important to note how AI, as seen from three closely related but widely differing attitudes about the current state of AI and its future potential, speaks to a story of technological progress (Stenvik, 2020). What I will argue is that these visions for humanity's future (i.e., what role AI technology will play in shaping the future) pose many interesting philosophical problems regarding among other things what a 'self' is, what the world is, and what the people inhabiting it are. As you shall see, I have come to label these three diverging scenarios as Pragmatic AI, and the pessimistic and optimistic futures of AI, and a more in-depth explanation of the future AI scenarios and the process of delimiting research, media, and books on AI technology into these three narratives is presented in further detail in Chapters 2, 3 and 4 respectively.

Through an analysis of how intelligence is conceived of in these different visions of the future of AI and its current state, one can begin to trace the outlines of a story about the future that has a direct influence on how we understand ourselves (Eisenstein, 2007; 2011; 2018). In the next section I briefly introduce a section of the debate around intelligence. Following this, I proceed to narrow the focus and scope of the thesis. I then conclude with the formulation of the research statement.

1.1. *A Brief Presentation of Intelligence*

In general, when speaking about intelligence most of us probably have at least some idea or understanding of what it means. We probably all know someone who we think is very intelligent, but if pressed to define what makes them intelligent, or how or in what way they are intelligent, one immediately finds that intelligence is not as easy to define as one might have thought. Smarts, cleverness, intellect, rationality, perception, IQ, brainpower, wit, brilliance, genius are just a few of the many synonyms related to intelligence. How, then, does one define intelligence?

In order to understand how artificial intelligence is conceived from differing points of view, we must first delve a bit into the concept of intelligence itself, although it should be noted

that this thesis is not about intelligence in itself, nor is it an attempt to establish whether computers or even humans can be intelligent or not.

Just as with artificial intelligence, the debate around intelligence itself as a concept is continually evolving (Goddard, 1946; Nisbett et al., 2012). In psychology and intelligence research, the debate revolves around different subjects such as what constitutes *general* intelligence, also known as the *g*-factor (Flynn, 2018), the degree to which genes and environment play a role in shaping or creating intelligence (Sternberg, 2013), whether IQ is a good measurement of intelligence, and the fairness of intelligence testing (Das, 2004), just to name a few. Chalmers (2016) writes that some intelligence researchers even question the assumption that there is such a thing as intelligence and that it can be measured: "Still, many others question these assumptions. Opponents hold that there is no such thing as intelligence, or at least that there is no single thing" (p. 48).

Goddard (1946) writes of the difficulty in settling on a single, satisfactory definition, and in a review of intelligence researchers' previous attempts at coming up with a good definition of intelligence settles on intelligence as being:

> the degree of availability of one's experiences for the solution of immediate problems and the anticipation of future ones. The man of a high degree of intelligence is the man who "Always knows what to do." Who is "never at a loss for a word." "Who always has the answer to the problem." (p. 68)

Goddard's common-sense definition, while perhaps not satisfactory from within the field of intelligence research, as we shall see later, closely mirrors some of the more modern definitions within the AI literature.

There are countless other definitions of intelligence to be found in all kinds of literature. Sternberg (2013) defines intelligence as being: "one's ability to learn from experience and to adapt to, shape, and select environments" (p. 501), where adaptation means being able to change in order to better fit with the environment one is in, shaping is to shape the environment to better fit oneself, and selection is the ability to choose a "new environment, often after adaptation and shaping have failed" (ibid.).

Tegmark (2018), when recounting the story of attending a Nobel Laureate symposium on artificial intelligence in which a panel of leading AI researchers failed to agree upon a definition of intelligence, jokingly writes: "there's no agreement on what intelligence is even among intelligent

intelligence researchers" (p. 49). He presents for his work a deliberately broad definition of intelligence so as to circumvent the debate around human versus non-human intelligence as well as different types of intelligence, thus placing it on a spectrum of ability and thereby avoiding the bickering over whether something is intelligent or not. His definition goes as follows: "intelligence = ability to accomplish complex goals" (p. 50). Bostrom's definition is also deliberately vague. In *Superintelligence* (2014), he writes that intelligence means "something like skill at prediction, planning and means-ends reasoning in general" (p. 107) allowing a hypothetical superintelligence to be intelligent while still possessing complete instrumental rationality that does not take any moral facts into consideration (ibid., p. 280).

1.2 *Focus and Scope of the Inquiry*

With the above in mind, it should again be noted that while it is an important underlying basis for the inquiry, the thesis is not about intelligence itself. It is rather a philosophical inquiry into three prevailing trends in artificial intelligence literature, and how, among other issues, the underlying assumptions about what intelligence is speaks to a story of technological progress that has deep roots in a particular mode of being.

Discussions about artificial intelligence (AI) have been around since at least the 1950s (Goertzel, 2007; 2013), but the debate about what constitutes "real AI" (Goertzel, 2007, p. 1162) and the potential consequences of the technology still abounds today (e.g., Bostrom, 2014; Kaplan, 2017; Madsen, 2019; Kulager, 2020). AI has the potential to drastically change the way humans work and do business together, and in some ways already has (Marr & Ward, 2019). Ultimately, other commentators note, AI may, or, depending on your view, will even with certainty radically alter the whole of society for the better through artificial general intelligence, or for worse through a badly programmed or manipulative superintelligence (e.g., Kurzweil, 2007; Bostrom, 2014). It might just turn out to be, as is often predicted, one of the most important technologies of the 21st century (Harvard Business Review, 2019; Kulager, 2019; Marr & Ward, 2019, p. 1-2).

The chosen perspectives represent three trends in the literature on AI in which the technology is either viewed as 'narrow' AI, a tool in the form of its business use case, which I take to represent the current state of the technology, or it is presented with a view to the future and what benefits or disadvantages the most powerful potential technology that mankind will ever create might bring with it. A fitting illustration of the two future thought currents appears in the paper "The Singularity: A philosophical analysis" (2016/2010), in which the philosopher David

Chalmers discusses the importance of taking the argument for the possibility of a Singularity[3] seriously. He writes:

> An intelligence explosion has enormous potential benefits: a cure for all known diseases, an end to poverty, extraordinary scientific advances, and much more. It also has enormous potential dangers: an end to the human race, an arms race of warring machines, the power to destroy the planet. (p. 37)

With such consequences at stake, it is not only important to discuss and understand what this technology is and is not, but also to discuss and analyze how these conflicting narratives about AI tell a very specific story not only about what intelligence is but also hint at a particular relation to technology. What is often neglected, however, is just how these prevailing narratives surrounding AI technology paint a picture of technological progress as part and parcel of human nature and the philosophical consequences of this particular framing of a more and more ubiquitous technology that has been touted as the most important technology of the 21st century.

1.3 *The 'Problem' of Artificial Intelligence*

First, according to Stenvik (2020), the stories about technological progress that originate from the technology sector are continuations of a very specific form of myth that contributes to a specific understanding of technology and the human-technology relationship. These technology companies at the forefront of AI research and development are not only enormously powerful economic actors, but they also shape most of, if not our entire digital infrastructure, and today, when larger and larger portions of many people's lives are spent in the digital sphere of existence in which AI is becoming increasingly more pervasive, if we do not have a fuller understanding of artificial intelligence, of the current status as well as the possible futures of the technology we might not have any say in what that future will look like. Second, at an immediate superficial glance

---

[3] Goertzel (2007) writes that the notion of a "Singularity" was introduced by the science-fiction writer Vernor Vinge who:

> ... used it to refer to the point at which scientific and technological progress occur so fast that, from the perspective of human cognition and perception, the rate of advancement is effectively infinite. The knowledge-advancement curve becomes effectively vertical, from a human perspective (p. 1164)

Oftentimes what causes the Singularity is an intelligence explosion as a result of the creation of the first human-level artificial intelligence. See section 4.1 for a more in-depth explanation of the 'intelligence explosion'.

or from a common-sense perspective, it might seem quite obvious what is meant by artificial intelligence, but a brief summary of AI literature from the above perspectives shows that intelligence, especially with regards to the artificial kind, is often defined in either a very broad or a very narrow sense. Bostrom (2014), in this situation, defines intelligence as the possession of "common sense and an effective ability to learn, reason, and plan to meet complex information-processing challenges across a wide range of natural and abstract domains" (p. 3). That does not necessarily have to be a problem, but as I shall show, this view has consequences in the form of shaping a particular perception of what it means to be intelligent that is seldom dealt with in the literature. Third, and perhaps most importantly, if AI, whether it be 'narrow' AI or AGI, will become the most important invention of the twenty-first century, it follows that the importance of looking at some of the dimensions left out of the discussion, the unquestioned assumptions of all sides of the debate, should not be understated. Moreover, seeing as the idea is of enormous practical as well as philosophical concern (Chalmers, 2016/2010, pp. 41-42) the idea should not be taken lightly. However, Chalmers approaches the subject of the possibility of a Singularity from an analytical philosophical angle, concluding that a singularity "is certainly not out of the question, and that the main obstacles are likely to be obstacles of motivation rather than obstacles of capacity" (p. 86). What he means by this is that the probability of a Singularity occurring at some time in human history is likely to be constrained not by technological, nor human intellectual capacities, but by constraints of incentive. That is, if we are aware of the rather sombre potential consequences of such a technology, will we actually want to create an intelligence far greater than ours?

Before presenting the research statement concluding this section some final limitations should first be clarified. Artificial intelligence, both narrow and general, touches upon many, many different related subjects and themes. The topic of consciousness, in this case, shall not be dealt with in further detail than what the authors representing the different outlooks on AI themselves do. Neither, shall I deal much more in-depth with intelligence itself as a concept. Instead, the focus shall be on how these stories as expounded in the three different future AI outlooks affect, as Geraci (2010) writes, "life on the ground". How do these stories shape the way we think of the future? What is excluded from the future in each scenario? How are they alike and how are they dissimilar?

I will argue in this thesis for the importance of, through works and concepts of Charles Eisenstein, conducting a philosophical inquiry into the shared assumptions about, among other issues, intelligence, the brain, and what stories of the future are implicitly told in these three

seemingly contradictory, yet closely related thought currents regarding possible future directions in of AI as well as the more pragmatic business use case for AI. For the purposes of analyzing the above, I have chosen the theoretical framework of Charles Eisenstein and his overarching concept of the Story of Separation and the related concepts of science, technology, reason, and control as the foundation for which the following research statement shall be examined:

> *A mapping of our contemporary understanding of artificial intelligence into three scenarios: the pragmatic business use case, the pessimistic, and the optimistic future of AI*

Some questions present themselves with regard to this research statement. The following questions related to the research statement are sought answered in the discussion and analysis:

- *What are the three scenarios? And how have they been arrived at?*
- *How is the intelligence of AI portrayed as intelligent in each scenario?*
- *How is the Story of Separation present in the Pragmatic, optimistic and pessimistic scenarios respectively?*
- *What role does AI play in the tale of technological progress in each scenario?*

1.4 *Structure of paper*

This thesis aims to contribute to the wider framework of the debate around artificial intelligence by challenging and critically engaging with the literature.

Chapter 1 has introduced the topic and context of the thesis, its relevance and importance as well as the research objectives. In Chapter 2, I begin by presenting the applied methodology, followed by an introduction of selected concepts from Charles Eisenstein's body of work and an argument for the usefulness of these concepts. The chapter concludes by clarifying the selection criteria for the included empirical material representative of the three AI scenarios. Through the concepts of Charles Eisenstein, Chapter 3 discusses and analyzes the current state of AI, which I have termed Pragmatic AI, after which the two future AI scenarios - Digital Utopia and Death by Banality, respectively - are presented and discussed in Chapter 4. Finally, Chapter 5 summarizes the central conclusions and addresses limitations and explores avenues for further research.

2. Analytical Framework

In this chapter, I present and discuss the methodological approach informing the discussion and analysis of the empirical material, followed by a discussion of the theoretical framework and concepts upon which the analytical framework is designed, ending with a brief presentation of the selection criteria for inclusions and omissions of data and literature chosen for the empirical material. The methodological approach is in part informed by Geraci (2010), and the concepts used to create the analytical framework are from the authorship of Eisenstein (2007, 2011, 2018).

2.1 *Applied Methodology*

In a footnote to the introduction of his book *Apocalyptic AI: Visions of heaven in robotics, artificial intelligence, and virtual reality* (2010), Geraci writes of the importance of relating to these apocalyptic elements present in popular science books on robotics and artificial intelligence on their own terms rather than seeking to figure out whether the predictions and convictions of the authors are right or wrong. In doing so, he addresses the importance of Husserl's concept of *epoché* in the history of religions, Geraci writes that...:

> The practice of *epoche* requires that we relinquish our presumption that we know what is true and what is not. In the study of foreign religions, this means assuming that the religious beliefs and practices of the object of one's study could be correct and efficacious. Rather than seeking to find "truth" or "falsity" in these beliefs and practices, one is better advised to seek out how they affect life "on the ground." *Epoche* applies equally to the promises made in pop science books. While it is not particularly valuable to either assent to or deny the futuristic promises of pop science books, as robotic and AI technology becomes increasingly prevalent in society, we would be well advised to sort out how those promises function within our culture, regardless of whether or not we accept them. (p. 167; emphasis in original)

Husserl's concept of *epoché* as conceived by Geraci does not seek to determine where 'truth' or 'falsity' are to be found, but how what is true and false within these beliefs (i.e., the stories that they tell themselves about themselves) affect the reality of the lived life. That is, the assumptions that make themselves felt in these works have, as Geraci notes, a function within our culture regardless of whether or not they are true, likely, or desirous.

Moreover, Geraci writes that Moravec and Kurzweil, two of the most famous exponents of what I describe as *optimistic AI* "take a dualistic approach to the world, one where physical and biological reality and bodily life are computationally inefficient and 'bad' while rational, mechanical minds and virtual reality are efficient and 'good'" (p. 2). Both Moravec and Kurzweil predict that humans "will upload [their] minds into machines and live forever in a virtual paradise" (ibid.). Not only does this situate Geraci within discussions of AI, but it also makes his adaptation of Husserlian concepts a fruitful point of departure for the methodology used in this thesis. This is, however, not to say that the concept of epoché, as used by Geraci, is applied in the analytical framework, but rather that Geraci's adaptation of Husserl's concept of epoché is informational for how I have approached the analytical framework as it relates to Eisenstein's concepts.

As such, to avoid recreating the aforementioned "dualist approach to the world," Geraci's application of Husserl is used as a twofold inspiration for the approach used in this thesis: On one hand, it informs the criteria from which literature and the empirical material that is discussed is selected and distributed under the three main thematizations of AI (i.e., pragmatic, optimistic, pessimistic), in that they are sought to be understood from their own premises. And on the other hand, it also informs the adaptation of concepts from Eisenstein, in that the concepts used from Eisenstein are developed from his approach to as best as possible engage with the analysis and discussions at hand (like Geraci uses Husserl's *epoché* concept). As such, the concepts are "excavations" from Eisenstein's authorship, but rather than striving towards being unquestionably loyal to those concepts (and thus creating the possibility of a dualism) they are applied to develop an analytical framework that will allow me to contribute to the discussion of AI in a way that Eisenstein himself might not be capable of.

Inspired by the methodology of Geraci, I shall use the overarching concept of the Story of Separation to examine how the story of artificial intelligence influences the present day. In most of Eisenstein's writings, he seeks to spell out the underlying mechanisms for how the Story of Separation has affected our entire being (2007; 2011; 2018). It affects our institutions, our science, our technology, and our relation to nature, our visions of the future both of business and ultimately of humanity itself. How is the Story of Separation with all its aspects present in the tales about artificial intelligence, and of what consequences are the framing of the visions of the future or cutting-edge current uses, these narratives of "the most important technology of the 21st century"?

2.2 *A Story of Technological Progress - The Philosophy of Charles Eisenstein*

Stenvik (2020) writes of one the tales of our times that has become so matter of fact that we have almost become blind to its very existence. Comparing them to magic spells, he notes that stories and our ability to imagine different futures might be one of the most unique things about human beings, yet there are stories so compelling that they take on a life of their own and shape the thoughts and actions of entire communities. One such story is the tale about technological progress.

This story of technological progress is one of the central themes in the philosophy of Charles Eisenstein. His overarching concept of the Story of Separation forms the groundwork for the theoretical structure that this thesis bases itself on. Through his explorations of the Story of Separation and its origins, Eisenstein traces how the Story of Separation has shaped the development of most, if not all, of modern civilization's most important institutions (Eisenstein, 2018, p. 8). This is the underlying story that runs in the very depths of our ontological and epistemological foundations and shapes our conceptions and understandings of the world as such.

Among the most conspicuous of these are what Eisenstein terms the Scientific and the Technological Program, through which the echoes of the Story of Separation reverberate within our stories of Self, the People, and the World. In Eisenstein's view, science and technology are not two separate, distinct realms. Within the Story of Separation, technology arises from science and is in turn the means for science to explore the depths and soaring heights of the universe as we know it (Eisenstein, 2007, p. 5).

It is because of these recurring themes of science and technology in his philosophy and the importance of stories in shaping our relation to the world, and our modes of being that the choice was made to use Eisenstein as the theoretical backdrop for this thesis. Seeing as Eisenstein writes so extensively on the hidden assumptions and shared agreements that we rarely question about ourselves, other people, and the world itself, the use of Eisenstein allows one to examine the subject of AI from a rather unconventional point of view. How is the Story of Separation and in particular the aspects relating to technology present in the representations and visions of the future that we are told AI will bring with it?

This section presents selected concepts related to the Story of Separation with particular focus on technology and science as conceived by Charles Eisenstein that will serve as the lens through which the three empirical AI scenarios, the Pragmatic, Death by Banality, and Digital Utopia, will be discussed. I begin by explaining the Story of Separation with an intention of creating a coherent, holistic analytical, theoretical framework from which follows an analysis of the

empirical material. The approach here is of both pragmatic and eclectic nature, in which key concepts from Eisenstein's collected works contribute to the analytical framework that is used to shed further light on aspects of artificial intelligence.

2.2.1 The Story of Separation

Who am 'I'? What is a 'self'? What is a human being? What is human nature? How do we understand and explain our world? Pervading most, if not all, of Eisenstein's writings is his search to answer these questions, or what could rather be called his attempts to untangle the webs of the Story of Separation. While these are not the specific questions that I set out to answer in this thesis, as shall become clear in the next chapter, these important questions in Eisenstein's philosophy are however closely linked to the subject/topic of AI and its surrounding narratives.

The Story of Separation is a way of understanding and critiquing the dualistic views of the self and ultimately the world we inhabit (Eisenstein, 2007, p. 145) and has had far-reaching effects on many aspects of our modern world, perhaps none more so than our modern institutions of science and technology. Therefore, the concept of the Separation, (i.e., humans separate from nature and from the world, and matter separate from spirit), allows for a reading of the empirical material in which the three AI narratives are viewed in terms of how they contribute to spell out a story of technological progress that is epitomized in the Story of Separation. It states that humans are separate individuals, separate from each other and from nature and is the defining myth of our civilization, the fundamental illusion of our society. In *Sacred economics* (2011), Eisenstein writes:

> Separation is not an ultimate reality, but a human projection, an ideology, a story. As in all cultures, our defining Story of the People has two deeply related parts: a Story of Self, and a Story of the World. The first is the discrete and separate self: a bubble of psychology, a skin-encapsulated soul, a biological phenotype driven by its genes to seek reproductive self-interest, a rational actor seeking economic self-interest, a physical observer of an objective universe, a mote of consciousness in a prison of flesh. The second is the story of Ascent: that humanity, starting from a state of ignorance and powerlessness, is harnessing the forces of nature and probing the secrets of the universe, moving inexorably toward our destiny of complete mastery over, and transcendence of, nature. (p. 1)

Here Eisenstein fleshes out the three most important aspects of the 'Story of Separation'. The 'Story of Separation' is a 'Story of the People' that consists of two parts: a 'Story of Self' and a 'Story of the World'. The second part, the Story of the World is a 'Story of Ascent' which Eisenstein first presents in *The Ascent of Humanity* (2007) and develops further in his following works. In the story of Ascent, it is humanity's destiny to "... transcend nature; to triumph over entropy, chaos, and decay; and to establish an ordered realm: scientific, rational, clean, controlled" (2011, p. 205). The story of Ascent states that the destiny of humans is to control, dominate, conquer, ascend above and ultimately fully transcend nature altogether (Eisenstein, 2007; 2011; 2018). It has two distinguishing aspects: the 'Scientific Program' that pursues a complete understanding of the world and the universe, and the 'Technological Program' that endeavors complete control over nature (Eisenstein, 2007, p. 5). While I shall also be discussing aspects of the Story of Self in parts of the analysis, the main focus shall be on Eisenstein's conception of technology, for it is the scientific program of complete knowledge that seeks "to convert reality into a data set" (Eisenstein, 2007, p. 502) and the technological program of full control that are most relevant for the purposes of this thesis. In the following sections I present Eisenstein's conception of science and technology as pertaining to the analysis and discussion of the ensuing empirical material.

2.2.2 Science

> … we imagine the light of science illuminating the few remaining mysteries of the universe, converting the unknown into the known, and subjecting the mysterious to the structures of human understanding and measurement. Consider though, that just as a campfire deepens the shadows beyond its circle of light, perhaps our science succeeds in illuminating only that which is within its purview, which we have deluded ourselves into thinking is the whole of reality, while making the vast beyond even more impenetrable. We have convinced ourselves that the world outside the campfire's circle does not even exist, or is not important, or will succumb to light as we build the fire higher and higher, consuming in the process every available bit of fuel.
>
> Charles Eisenstein, (2007). *The Ascent of Humanity*, p. 45.

Through his concept of the Story of Separation, Eisenstein describes an attitude to science that is perhaps best epitomized by a common belief often expressed in sentences such as "science does not have all the answers… yet," or something to that effect. There is an implicit assumption in

this statement that complete objective knowledge is possible, that there is such a thing as an objective reality "out there" (Eisenstein, 2007, p. 329) and that science has access to this realm of objective reality without affecting it, thus preserving its objectivity and splitting the realms of reality into a dualism of observer and reality. Eisenstein argues that this is an attitude that pervades throughout the scientific community, even though philosophers of science might object to this claim, stating that full knowledge and total control of the universe is highly unlikely (2007, p. 5; 2018, p. 246).

The point Eisenstein makes is not that science is of no value in determining truth, or that we can not trust its results, nor that we should do away with science altogether. It is rather to highlight how the "metaphysical assumptions" and "institutional expressions" of science springing from the Story of Separation have reduced reality to quantifiable bits of information, matter into "generic particles" and nature into homogenized, substitutable goods (Eisenstein, 2018, p. 249). If science maintains "our culture's main map of reality" (ibid., p. 245), then it is deserving of our scrutiny. Science, as it stands for many people today, determines what is real, what is possible, and therefore it is important to ask how science as an ideology, as an outgrowth of the Story of Separation, following rather than preceding our understanding of self and universe (2007, p. 115), limits the kinds of technology that are conceivable from within this particular, separate, dualistic view of the world (2011, p. 248).

Furthermore, throughout his writings, Eisenstein points out how modern science has in many ways become similar to a religion. He writes that:

> … science has become the new religion, complete with its story of cosmogenesis, its mysterious explanations of the workings of the world couched in arcane language, its priests and their interpreters, its hierarchy, its initiation rituals (the PhD defense, for example), its system of values, and much more. (2011, p. 167)

Science tells us the story of how everything - both the universe, the Earth, humans, and all living things - came into being as well as how they function. It tells us these stories in an esoteric, impenetrable language that only the initiated (i.e., the scientists) understand and are able to communicate and disseminate to the masses. The priests of science only become ordained once they have studied the orthodox holy scriptures (i.e., the corresponding scientific canon/precepts) for a certain number of predetermined years and have passed the right examinations, etc.

This is not an attempt by Eisenstein to dismiss science altogether. The Scientific Method itself, writes Eisenstein, "rests upon a beautiful impulse, an ideal of humility and intellectual non-attachment that would serve any system of belief in good stead" (2007, p. 114). His point is rather to illustrate the inherent constraints and limitations of science and the scientific worldview, and thus open up new realms of the possible within science itself. In *Climate*, he writes that his call is "not to discard science but to expand it, to include what it has ignored" (2018, pp. 248-249). If we then think of science expanding or opening up so as to include what it has hitherto ignored or excluded, what could it become? For example, what might science (and in its wake, technology) become, Eisenstein asks, if it were to refrain from "privileging quantitative over qualitative reasoning" (ibid., p. 248)?

Inseparable from Eisenstein's discussion of science and technology is the concept of 'reason' or 'rationality'. The realm of pure rationality, where decisions are made free from "our biological heritage of instinct and emotion" (Eisenstein, 2007, p. 116) is the realm of total certainty in which doubt is but a symptom of not having taken into account all the necessary data points required to make a reasonable decision. However, just as with science and the Scientific Method, Eisenstein does not recommend the eradication of reason. Reason has its purpose, its time, and place. It "is still a valid and useful tool; it is only when it becomes a reflexive habit rather than a conscious instrument that it is limiting" (ibid., p. 329). The problem for Eisenstein is thus when reason alone is elevated to the supreme faculty of the mind, to the highest cognitive function that humans are capable of, because this relegates all other forms of cognition to a subordinate level, and, seeing as humans are above nature, also to a sub-human level. Reason alone is the capacity that sets us apart from other animals and nature, but its relatively narrow mode of cognition (ibid., p. 330) cannot solve all of our problems. Making a rational decision means that it is made without recourse to emotional attachment (ibid., p. 114), but "... we delude ourselves when we cite logical reasons, which are actually rationalizations or justifications for our decisions" (2007, p. 53). But what if reason is not the only alternative? What if emotions, intuition, instinct all have a role to play? What if other realms of cognition realize a function that, because we have elevated reason to the supreme faculty, the use of which we can no longer recognize?

Aside from science, the most important concept for the purposes of the subsequent discussion is that of technology which I shall now present.

2.2.3 Technology

> Like an addict trying to hold his life together, we shift debts, create rationalizations, and generate long-term consequences to solve short-term problems, pretending all the while that everything is under control. "Science will find the solution," we think, as we manage problems by putting them off until a future day of reckoning.
>
> Charles Eisenstein, (2007). *The Ascent of Humanity*, p. 506.

What if the techno-optimists are right that any problem "is just a technological hurdle in humanity's race to its glorious destiny" Eisenstein asks (2018, p. 160)? But what if that is not the right way to frame the question? Eisenstein implores his readers to instead consider: what kind of world do they want to live in (ibid., p. 163)? What do we take with us into the future and what do we leave behind? This section presents Eisenstein's concept of technology as it appears in the books *The ascent of humanity* (2007), *Sacred economics* (2011), and *Climate* (2018).

Technology is the means to manipulate nature, "driven by the urge for comfort and security; that is, the avoidance of pain and the insurance of survival" (Eisenstein, 2007, p. 278). It is "one of the key 'hormones' of the human metaorganism" (Eisenstein, 2011, p. 252). Just as science, the Scientific Method, and reason or rationality, technology is not a mistake, it is part of being human. "There is no such thing," he writes "as a pretechnological human being" (2007, p. 38). Before the technologies of separation, before the industrial revolution, even before agriculture we had stone tools, before those we had fire, and even before fire we had language and culture. The distinction between animal and human technology is often based on humans being the only animal using tools to make more tools. This speaks to a cumulative effect where technology, being a natural extension of an animal's body, so to speak, and therefore not a "mistake", is rather an acceleration of a development that has been a long time underway (ibid., p. 39).

There are two main parts to Eisenstein's conception of technology: The Technological Program and the technological fix. The definition of Technological Program is to perfect human control of nature (2018, p. 121), meaning in part that through reducing reality to data (2007, p. 506), we can quantify all the causal relationships in a complicated natural system, after which surely we can also improve upon the random processes of nature (Eisenstein, 2018), p. 121).

The Technological Program is the all-important myth of the *Ascent of Humanity* and it tells us that we can indefinitely postpone the solutions to our problems because one day we will find a technological fix that will cure all our ills and problems (2007, p. 414). Eisenstein's attitude to the technological fix, the indefinite postponement of the future, so to speak, is perhaps best illustrated

in the following quote where he argues that it is not possible to entirely bring to completion the Technological Program, but that certain problems undoubtedly do give in to our "methods of control, the technological fix. Looked at piecemeal, the Technological Program is a great success. We have attained to a realm of magic and miracles. Godlike powers are ours. Yet somehow, the world around us falls apart." (2007, p. 479).

The "ideology of the technological fix" sees technology as the cure for "the problems caused by existing technology" (Eisenstein, 2011, p. 161), meaning that "technology creates problems that necessitate yet more technology" (ibid., p. 252). Technology thus becomes an over-complication of already complicated systems. Each incremental fix brings with it ever-increasing complexity until finally, the system collapses under its own weight (2018, p. 173). The technological fix is epitomized by the attitude of "techno-optimism" (Eisenstein, 2018, p. 55) in which human ingenuity and creativity is seen as limitless so as no matter what obstacles we are faced with, we shall always find a way to overcome them through the technological fix (ibid., p. 168).

Central to Eisenstein's notion of technology, and hence separation, is the concept of control. In a recent essay on the topic of the Coronavirus, he writes of "the Reflex of Control" in which he describes how our societies'

> ... go-to crisis responses, all of which are some version of control, aren't very effective in addressing these conditions [of world hunger, addiction, autoimmunity, suicide, or ecological collapse]. Now along comes a contagious epidemic, and finally we can spring into action. It is a crisis for which control works: quarantines, lockdowns, isolation, hand-washing; control of movement, control of information, control of our bodies.

While in this context he is writing on the Coronavirus, the Reflex of Control is nothing new. It springs from the Story of Separation through which we have become "deeply habituated to thinking that a solution means more control, more detailed measurement of all the variables, a more comprehensive design" (2007, p. 408). Our impulse to control is the go-to problem-solving mode when faced with crises, but not all aspects of existence are well suited for control. Increased control does not necessarily make us safer, more comfortable, more convenient. It does however uphold the impulse for even more control (Eisenstein, 2007, p. 319). But does not technology fulfill our true human needs? Through the example of telecommunications, Eisenstein shows that while technology does indeed fulfill some human needs, allowing us to "stay in contact with people with whom we share close emotional and economic ties" (p. 80), during most of human existence,

these people would usually have been close enough, so as not to need the technology in the first place. Thus, the technology grew only to "meet a need [...] when other developments in technology and culture spread human beings farther apart and splintered extended families and local communities" (ibid.).

Technology as we know it today is the apex of our impulse to control, to ascend nature altogether, but, Eisenstein asks, what might technology look like if it were instead rooted, not in the Story of Separation, but in another story, one of interbeing, in which humanity's purpose is not to transcend nature, but to make the world more beautiful and to restore the health of the planet (2011, p. 393)?

2.2.4 The Story of Interbeing - A Final Remark

Up until now the Story of Separation paints a pretty bleak picture of our current understanding of the world, the universe and ourselves, but Eisenstein also presents an alternative to the Story of Separation. While the Story of Interbeing does figure in the following discussion, it is not included here for the express purpose of functioning as an analytical tool, but for giving a more whole picture of Eisenstein's philosophy and to illustrate that he not only seeks to critique and challenge our current stories and myths of the world, but to also present an alternative.

While the 'Story of Interbeing' is the obverse of the Story of Separation, it goes beyond common tropes like "we are all one," or "all we need is love," but that does not mean that love and interconnectedness or interdependency do not play a part in the Story of Interbeing. Eisenstein writes that his own "very being" takes part not only in the being of you and me, but also that of the forests, the oceans, the whales, and the elephants (2018, p. 22). It is this expansion of the 'self' that encapsulates Eisenstein's Story of Interbeing. If we are not separate, atomistic individuals, then what happens to you also happens at some level to me (ibid.). We can no longer isolate what happens to the Earth and the rest of her beings to another independent variable in an equation in which we have calculated the effects of any given action on all other independent variables within a fully quantified system. This form of linear cause-and-effect thinking leads to a trap in which we have expanded a linear domain onto a "hopelessly nonlinear" real-world system containing complex living organisms (Eisenstein, 2007, p. 134) that we in our linear mode of thinking reduce to

Eisenstein asks us instead to consider that we might not be able to fully calculate the effects of our actions in a direct causal way. If, from the perspective of the Story of Interbeing, what

happens to me also happens to some extent to you, then the stories we inhabit also need to be factored into our accounts of the effects of our actions. But this too would be to fall into the trap of linear thinking. Thinking that we could "account" for "effects" speaks to the same mode of thinking in which we can calculate the direct causal relationship.

If, as Stenvik (2020) argues, stories are not only important for our ability to imagine different futures, but are also a tool for understanding, affecting, and shaping the stories of our future, then the philosophy of Charles Eisenstein is a philosophy of the future, and the importance of illuminating the Story of Separation as well as the telling of multifarious stories of the future can not be overstated.

Finally, as I have now shown, the concepts that make up the Story of Separation are closely related and interdependent. As such, it is not possible to separate Eisenstein's notion of technology and control from his concepts of science and reason. How, then, does the story of separation appear in the three AI scenarios? The concepts described in the above are the key concepts that are needed to engage with the ensuing analysis and discussion. However, this is not a fully exhaustive presentation and discussion of all relevant concepts, and when new concepts are introduced, they are explained to the extent needed for the specific discussion.

Before proceeding to the analysis and discussion in Chapters 3 and 4, I shall first briefly present the selection criteria for the empirical foundations for the analysis.

2.3 *From Inductive to Deductive (and Back) - Selection Criteria for the Empirical Material*

This section describes how the three AI scenarios that make up the empirical material of this thesis were arrived at. In short, through both informal conversations with professionals who work with AI on a daily basis (C. B. Nielsen, personal communication, December 16, 2019), podcast episodes (e.g., Fridman, 2020a; 2020b) and reading peer-reviewed journal articles, books, and news articles on just about anything on the topic of AI,[4] it was possible to trace out certain trends regarding the topic of AI. It was through these readings that I settled upon the three common themes or underlying scenarios as I term them.

[4] I.e., philosophy of AI (e.g., Dreyfus, 1974), economics of AI (e.g., Brynjolfsson & McAfee, 2018; Ernst, Merola & Samaan, 2019), intelligence theory (e.g., Flynn, 2006), ethics of AI (e.g. Bostrom, 2003; Crawford, 2013), AI and data (e.g., Harari, 2017; First, 2018), algorithmic bias (Manyika, Silberg & Presten, 2019), decision-making (e.g., Brighton & Gigerenzer, 2015) etc.

The term 'Pragmatic AI' is inspired by a passage from Goertzel (2007) in which he describes how much of the (then) current research on AI focuses on "more narrowly-defined research directions that have the benefit of far more easily leading to scientifically demonstrable and/or pragmatically useful results" (p. 1163) as opposed to focusing on the more speculative, theoretical, albeit substantial, benefits of AGI. For the Pragmatic outlook on AI, a plethora of business and consultancy reports on the subject of AI, analytics, and big data and from this the recurring themes presented in Chapters 3 and 4 were gathered.

The pessimistic and the optimistic future scenarios are inspired in part by Bostrom (2014) and Kurzweil (1999, 2007, 2008) respectively, however the terms "Death by Banality" and "Digital Utopia" are taken from Tegmark (2018) in which the author describes different hypothetical end scenarios on the spectrum of possibilities regarding what happens after a true AGI is invented (pp. 161-202). Apart from settling on the themes or scenarios as they appeared to me, one of the primary selection factors when choosing representatives for the pessimistic and optimistic future scenarios of AI, respectively, was a certain form of cross-referencing within the literature. Except for Tegmark (2018) who himself refers to all the other authors of the selected texts for these scenarios, and Chalmers (2016) who does not refer to Goertzel in this specific article, the authors include references to at least one work of each of the other authors selected as representatives of these views.

The authors I have chosen to represent these trends range across a spectrum of positions or attitudes regarding what we can expect if or when the first true AGI is invented. This spectrum ranges from the overly optimistic to the emphatically pessimistic. See also Figure 1 in Chapter 3 for a rough illustration of each author's position on the matter. Kurzweil, for instance, ranges on the far end of the optimistic side of the spectrum, while Bostrom ranges on the far end of the pessimistic side, and Tegmark (2018) places somewhere around the middle with a more neutral attitude to the subject.

These scenarios appear in numerous books on the subject as well as peer-reviewed articles in scientific journals and electronic and print media articles (E.g., Kurzweil, 1999, 2005, 2008; Goertzel, 2007, 2013; Chalmers, 2016; Bostrom, 2014; Harari, 2016; and Tegmark, 2018). As shall become clear, some of the authors selected as representatives for these visions of AI lean predominantly to one of two sides on the debate around what the technology of AI will bring to the future of humanity. I have previously described the two sides as consisting of two opposing views on exactly how AGI or true AI is going to impact humans. On the pessimistic side, proponents tend to view AI as the most potentially dangerous technology in existence today. The

other side, the optimists, view the creation of AGI as humanity's top priority, the most important technology of the twenty-first century because of the potential to solve "hard problems and explore various positive avenues" (Goertzel, 2013, p. 130) like eradicating poverty and curing all disease (Chalmers, 2016; Tegmark, 2018).

Just as with the Pragmatic attitude to AI technology, by choosing authors with wide-ranging viewpoints across the spectrum of optimistic and pessimistic outlooks on what AI technology will bring to the future, I hope to paint a fuller picture of the debate around the technology itself. Kurzweil and Bostrom's inclusion was settled upon because of their respective attitudes to the perceived positive or negative effects of future AI. Alternatively, I could have chosen to use only Kurzweil for the optimistic outlook and Bostrom for the pessimistic, but this would have led to a narrower range of opinions on the matter and a thinner description. The inclusion of a broad spectrum of opinions makes for a more interesting analysis as it relates to how artificial intelligence relates to a Story of the People (Eisenstein, 2011, p. 1), allowing us to look at similarities and divergences while gradually contributing to a picture of the Story of Self and the World that is being painted along the way.

While Tegmark (2018) is not necessarily an adherent to the trans- or posthumanist view that Kurzweil, Bostrom, or other proponents of the Singularity idea are (e.g., Goertzel, 2007; Geraci, 2010; Chalmers, 2016), his more descriptive approach, presenting different objections to what a future AGI scenario might look like or how humanity might evolve into a human-machine hybrid, what he terms "Life 3.0", does warrant his inclusion. With Life 3.0 the artificially created life forms are able to design not only their own software but also their own hardware and is made possible when humanity manages to create an AGI enabling us to accomplish virtually any goal and (Tegmark, 2018, pp. 29-30).

In Goertzel (2007), the author discusses the concept of artificial general intelligence as opposed to other variations on the term such as "strong" AI and human-level or human-like artificial intelligence. The important thing to extract from this is that no matter if one prefers "strong" AI, human-level, or AGI, it needs to be seen in relation to or opposition to "narrow" AI. All terms are problematic within their own rights. Goertzel (2013) writes that even though many different technologies will play a role in "the intelligence explosion" giving humans new and radically posthuman characteristics, artificial general intelligence will play a "special role" (p. 129) as the essential or fundamental technology serving as a catalyst for this radical change.

Chalmers, as I have already discussed, approaches the subject from an analytical angle, in which he argues that given the likelihood of an intelligence explosion occurring at some point,

however small that probability may be, we must take the idea seriously. The very possibility of the Singularity, and the great potential benefits as well as enormous negative consequences that come with it, warrant philosophical scrutiny of the idea (Chalmers, 2016).

Although Harari (2016) also writes on the topic of the eventuality of machines and algorithms gradually replacing human beings, the decision not to include Harari as one of the representatives of this particular trend in the literature was made based on the fact that he approaches the subject through an historical analysis of human societies tracing in various points of depth the development from hunter-gatherer societies to agricultural, to industrial to liberal societies, leading to a less singular focus on the topic of AI itself.

As such, this overall combination of methodology, theory and empirical selection criteria leads to each chapter being a representation of the multiplicities within the selected literature, to catch both dispersions and tensions "within" each separate direction, enabling a discussion of each "on its own terms". Using the analytical framework developed in this chapter, I discuss each scenario in turn in chapters 3 and 4 respectively.

3. Current AI - The Business Use Case

This section presents first the pragmatic, thereafter the pessimistic, and finally, the optimistic AI scenarios and analyses each accordingly through the lens of Charles Eisenstein's overarching concept of the Story of Separation the analysis seeks to show how each imaginary portrays intelligence, human nature and ultimately continues in telling the Story of Separation. I begin by presenting what I have termed Pragmatic AI, followed by the Digital Utopian perspective, and conclude with the Apocalyptic 'Death by Banality' scenario.

The three following scenarios represent, as I have explained in section 2.3, three currents of thoughts within both the general discussions and the scientific literature on the subject of AI (see, e.g., Table 1 below). The trends reflect predominant, but not exclusive, attitudes towards the technology of AI, and are discussed using Eisenstein's concept of the Story of Separation and the related concepts of technology, science, reason and control.

	Pragmatic AI - The Business Use Case	**Future AI - Death by Banality or Digital Utopia?**
Optimistic	Davenport, 2018; Marr and Ward, 2019	Kurzweil, 1999, 2008; Goertzel, 2007; Chalmers, 2016/2010
Neutral	Fleming, 2019; Madsen, 2019	Harari, 2016; Tegmark, 2018
Pessimistic	Kulager, 2020	Bostrom, 2014

Table 1: Rough sketch of the selected authors' relative attitudes on the optimism/pessimism scale regarding their views on the current state and future of AI.

3.1 *Pragmatic AI*

> ... there is a massive amount of hype and confusion about AI. Some see it as the ultimate threat to our civilization, while others believe AI is the savior that's going to solve humanity's biggest challenges, from tackling climate change to curing cancer. The aim of this book is to cut through the hype and scaremongering and provide a cutting-edge picture of how AI is actually being used by businesses today.
>
> Bernard Marr & Matt Ward (2019). Artificial intelligence in practice: How 50 companies used artificial intelligence to solve problems, p. 1.

This section discusses some of the more pragmatic approaches to AI and presents some of the key challenges that businesses are faced with in regards to applied artificial intelligence (Davenport, 2018, pp. 171-197; Marr & Ward, 2019, pp. 325-326). I begin by discussing narrow AI as conceived of in Davenport (2018) and Marr and Ward (2019) and various consultancy reports and peer-reviewed journal articles. What is it? How does it augment human decision-making and contribute to new knowledge? What is the context in which it is used, that is, how do businesses actually use it to help them achieve their goals?

By the Pragmatic scenario, I refer essentially to what can also be called the business case for AI. In this sense, AI is all about businesses becoming data-driven and enhancing decision-making through functional or narrow AI and machine learning, and many of the overly hyped or pessimistic predictions about what AI will bring in the future are severely overrated (Kaplan, 2017; Madsen, 2019). Digital businesses such as Amazon, Facebook and Google generate large amounts of data, and in order to make sense of it and create better insights into their own business or their customers' habits, a number of AI-based technologies can show them how to make better decisions based on facts. A number of consulting reports from many of the world's largest consulting firms such as Boston Consulting Group or McKinsey & Company (see, e.g., Gourévitch et al., 2017; Harrison & O'Neill 2017; Henke & Kaka, 2018), as well as articles in popular business magazines such as Harvard Business Review (see, e.g., Mahidar & Davenport, 2018) speak of the need for businesses and organizations to fully utilize the potential of the massive amounts of data they have available to them in order to help "improve the speed and quality of decision-making" (Smith et al., 2019). In this scenario, the dichotomy of a general AI on one hand or a destructive superintelligence that overtakes humanity on the other is not a matter of concern.

Davenport (2018) acknowledges that the potential of AI is enormous when he writes that: "In the short run, AI will provide evolutionary benefits; in the long run it is likely to be

revolutionary" (p. 7). Marr and Ward (2019) write several times of an "AI goldrush" (e.g., pp. 2 & 323), and a fourth industrial revolution "that is going to transform all parts of business and society" (p. 3). Seen from this perspective, what is important is not the potential technological developments of an AGI or how futurists and researchers think AI is going to affect society over the next 50 years, but how companies use AI in a practical sense in order to be better at conducting business, and, if they are proficient at it, ultimately gaining a competitive advantage over their competitors.

Before questioning and discussing the potential tensions between AI as beneficial for businesses and consumers, as a phenomenon of possibly significant societal (or even existential) influence, the relevant distinctions and differences between narrow AI and AGI must first be properly contextualized.

3.1.1 What Is Narrow AI?

> Despite the progress, many hard problems remain that will require more scientific breakthroughs. So far, most of the progress has been in what is often referred to as "narrow AI"— where machine-learning techniques are being developed to solve specific problems, for example in natural language processing.
>
> James Manyika et al. (2018) The promise and challenge of the Age of Artificial Intelligence, p. 2

"Narrow" AI is a term often used in opposition to AGI. The most common example used to illustrate narrow AI is that of a program, such as chess-playing Deep Blue (Tegmark, 2018, p. 30) or the more recent Go-playing AlphaGo (Wang et al., 2016). Such an "intelligent" system is only intelligent at playing a specific game and its knowledge is not directly transferable to other cognitive spheres[5]. The term describes an AI that "performs one narrow task, as opposed to artificial general intelligence, or AGI, which seeks to be able to perform any intellectual task that a human can do" (Henke & Kaka, 2018, p. 20). In a business setting, the focus is often on narrow AI because of its "near-term business potential" (ibid).

Davenport (2018) writes that it will be "dangerous to do nothing in [the AI] area, or to move too slowly" (p. 8). He is, however, not writing of the existential dangers of a superintelligence

[5] Although the creation and application of narrow AI for games like chess or go in itself is only relevant to apply within very specific frameworks, the process of creating an AI that can learn itself to play games as complex as chess and go remains transferable to much wider spheres (Wang et al., 2016).

or the dangers of losing out on the benefits of AGI, but instead of businesses being left behind other businesses that have managed to "[realize] the power of this technology" (ibid.). What really counts is how narrow AI helps businesses get ahead of their competitors, and there are plenty of examples of how all kinds of companies are doing just that. The near-term business potential of narrow AI has not only the potential to generate enormous improvements in business performance in all sectors of the economy and across multiple business functions (Manyika et al., 2018, p. 3). Bughin et al. (2017) note several ways in which AI helps businesses increase productivity. It can be used to continually optimize "assets and processes," assemble the finest teams of "people and robots," boost "quality and reliability," and "prevent downtime for maintenance" (p. 26).

Davenport moreover highlights the importance of noting that our understanding of the "day-to-day application" (2018, p. 10) of artificial intelligence is filled with ambiguities. It is important, he notes, that narrow AI should not be conceived of as a single coherent whole, as one thing in particular. It is a host of different technologies that are colloquially known as AI. He lists seven key technologies that encompass the entire subsets of technologies that make up AI. These are statistical machine learning, neural networks, deep learning, natural language processing (NLP), rule-based expert systems, physical robots, and robotic process automation (ibid., pp. 11-17). And though providing a detailed overview and full contexts to these seven key technologies is not within the scope of this thesis, it is illustrative of the fact that even "narrow" AI is remarkably multifaceted and relevant when referring to what these individual technologies, collectively known as AI, render possible for modern businesses.

According to Marr and Ward (2019), the key uses of AI to businesses are threefold: First, they can change the way businesses "understand and interact with customers," second, they can "offer more intelligent products and services," and third, they can "improve and automate business processes" (p. 6). Davenport (2018) shares this point of view, writing that AI assists businesses in three key activities, allowing them to automate "structured and repetitive work processes," and gain insights "through extensive analysis of structured data" and engage "with customers and employees" in new ways utilizing, for example, NLP chatbots (p. 41).

For businesses to actually benefit from implementing AI in the above manner, they must first make sure that their business models are up-to-date and fitting with its strategic goals (Harvard Business Review, 2019, p. 11; Marr and Ward, 2019, p. 323; Smith et al., 2019). This involves investing in creating thorough and detailed AI and data strategies, identifying crucial business opportunities, and challenges that can benefit from the application of AI. In addition to this, the culture of the organization must become data-driven in order to become a catalyst for developing

an organization in which humans and machines work together to bring substantial business results (Smith et al., 2019; Marr and Ward, 2019). Developing and implementing the correct and corresponding AI and data strategies, as well as establishing a data-driven culture are not the only hurdles that businesses must overcome to fully reap the benefits that AI can bring with it. The next steps include ensuring that management and leadership are fully invested in the push to become a data-driven organization as well as making sure the right technological capacities are in place, and maybe most critical of all, the organization must ensure that a high-quality data infrastructure is established providing the business (Bughin et al., 2017).

While Davenport (2018) goes more into depth with how different aspects of the technology actually work, Marr and Ward write almost exclusively about how and where AI is actually being applied in businesses. For the purposes of this chapter of the thesis it is sufficient to understand that, as long as businesses want to improve, learn, and gain new understandings and knowledge, the possibilities of narrow AI within a business context are virtually endless (Marr & Ward, 2019). Equally important within the context of Pragmatic AI, the current business use case of AI, is a discussion of how AI can contribute to better decision-making and new knowledge.

3.1.2 What to Do with All the Data - To Measure or not to Measure?

Modern businesses collect, create, and have access to more data than ever before (Bughin et al., 2018; Domo, 2019), and as presented in the introduction the number of bits of data being created at any given moment is astronomical. Therefore, with all the data at their hands, businesses have an untapped resource just waiting to be turned into something useful. Now, data is not just data. If data quality is not of the highest possible caliber, the insights gleaned from the AI algorithms can be fraught with error (Henke & Kaka, 2018). Data quality is perhaps the most important factor if a business wants to fully benefit from AI. Ransbotham et al. (2017) write that: "Sophisticated algorithms can sometimes overcome limited data if its quality is high, but bad data is simply paralyzing" (p. 8).

This shows how the importance of high-quality data can not be overstated. Therefore, the first step[6] has the tripartite requirements of ensuring not only the quality of the data but also that sufficient quantity of data and the right data infrastructure are in place as well (Harvard Business

[6] After ensuring the correct AI and data strategies are in place and a data-driven analytics culture is established. These are costly affairs, I know. One could almost speculate about why the major consultancies make such an effort to position themselves as experts in AI and data strategy and culture.

Review, 2019; Ransbotham et al., 2017). Moreover, the data collection, processing, and storage infrastructure must be in line with the AI and data strategy of the organization.

Once the accuracy (Silberg & Manyika, 2019), relevance (Bughin et al., 2019), and consistency (Ransbotham et al., 2017) of the data has been assured, one can then assume that the algorithmic output (i.e., predictions and decisions) can be trusted. However, this does not take into account algorithmic aversion (Burton, Stein & Jensen, 2019), which states that humans are oftentimes averse to use algorithmically enhanced decision-making aides, nor the need for explainability, that is that decision-makers are able to understand how an AI has come up with a recommendation, prediction or result and can further communicate this to the relevant parties (Silberg & Manyika, 2019). Nor does this take into account various types of bias that might play a role in affecting the objectivity and reliability of the algorithmic output.

The types of biases that might present themselves are widespread, as demonstrated by such diverse examples as sample bias, embedded bias, measurement bias and fairness of data bias[7]. Sample bias refers to bias implicit in the data that an algorithm is trained on. For example, in 2015 Amazon had to discard an AI-powered recruiting tool that heavily favored resumés from male applicants over those from women applicants (Dastin, 2018). Because the model that Amazon used had been trained on resumes received over the preceding 10 years, and because most of those applicants had been male, the model reflected that in its "preference" for male applicants.

Embedded bias refers to human bias and the way that the algorithm is programmed or the models are designed. A human being might implicitly program his or her own biases into the model. Often, the data that is used to train an AI model is tagged by humans that then directly train the algorithm with their implicit biases (Satell & Sutton, 2019).

Measurement bias is the kinds of bias in which the bias is inherent to the data collection method itself. A classic example of this is the app *Street Bump* developed by the City of Boston in 2011 in order to detect potholes through GPS and accelerometer data collected from citizens who had the app installed on their smartphones. The surprising result was that a larger share of the potholes was found to be in relatively more affluent areas of the city. This is because residents from lower-income areas as well as older residents were less likely to have a smartphone in the first place, therefore biasing the data from the outset of the collection toward collecting more and

[7] There are many more biases, both algorithmic and human, but an exhaustive review of them all is beyond the scope of this thesis. The point of including these biases is to demonstrate the diversity and contexts of biases in the use of AI.

better data from the affluent neighborhoods where smartphones were more prevalent (Crawford, 2013; Satell & Sutton, 2019).

The final example of a bias to be overcome is that of the fairness of the data itself. While the data might reflect the historical reality of the population, the decisions made on the basis of the data might still not be based on fair criteria. The example of algorithms used to combat crime serves to illustrate this point. Various police departments in the United States use predictive algorithms to help them "strategize about where to send their ranks" (Hao, 2019). They also use facial recognition systems to help single out suspects, but these algorithms have repeatedly been shown to be particularly inept at recognizing the faces of dark-skinned persons. Lastly, there are the risk assessment tools that are used to predict whether an offender is likely to commit another crime in the future (Davenport, 2018, p. 176; Hao, 2019). These algorithms are often trained on historical crime data, and this will inevitably carry inherent biases of the criminal and judicial systems over into the predictive algorithmic systems (ibid.).

As an extension of these biases, explainability of an algorithm's predictions or decision making remains a problem. Davenport (2018) also writes of the trade-off between transparency and performance. In some cases it is not all that important how the AI arrives at its decisions or recommendations, and what is important in these cases is that it works and the predictions are accurate, or the decisions made are sound. But in some cases transparency is very important. In financial services, regulators demand that it is possible to explain the inner workings of an algorithm. Moreover, in the GDPR laws that came into effect in 2018, it is stated that "individuals affected by computer-based decisions have a right to know why and how the decision was made" (ibid., p. 181).

This makes for a dilemma in relation to the transparency/performance trade-off. If an algorithm works perfectly well, yet is difficult or near impossible to explain with regards to its criteria of validity or biases, then how can businesses solve the problem of granting an individual their right to know why and how a decision was made? Explaining how algorithms arrive at their decisions is quite complicated, and AI technologies like deep learning "make it virtually impossible to know what features or variables in the model mean, what impact they had on the outcome, and how the model arrived at an outcome" (ibid., pp. 179-180). This leads to what is commonly known as the "black box" problem in which all we can say about the technology is that it works (relative to the human defined goals set out from the start), but we can not explain how or why it works, nor interpret its results in any meaningful manner (ibid., p. 178).

To sum up, businesses, governmental institutions and other types of organisations have enormous amounts of data at their disposal. However, the data must pass several quality checks before it can serve as foundation for an AI algorithm that can create novel insights. Insights that can be used by businesses, institutions and organizations to better understand and interact with their employees, customers and users; research and develop more intelligent products and services; or improve and automate different business processes. There are therefore many pitfalls through which organisations that wish to fully utilize the enormous potential of AI must navigate through before they can actually reap the benefits.

Now, if only we could avoid bias altogether, then we could achieve absolute certainty. The search for objectivity can finally be called to a halt, because with big data, cloud computing and ultra-fast artificial intelligence algorithms, the numbers speak for themselves (Crawford, 2013). But assumptions that numbers speak for themselves or that data can be objective seems futile at best with aforementioned biases in mind. As I have shown, there are many ways that data can be biased, and the examples listed above are by no means all-encompassing. The search for complete objective knowledge through more and more extensive use of data and number and ever more complicated machine learning algorithms and other AI tools, is in Eisenstein's view a way of controlling the world through extending what can be measured to more and more domains of existence (2007, p. 61-62). If viewed through Eisenstein, the proliferation of data in business and the quest to utilize it for the betterment of business performance speaks into the logic of ever more quantification of more and more aspects of the world. If we have the pure, true facts, the perfectly cleansed high-quality data and all the numbers necessary, we can then make the utmost rational decisions. This is how we achieve certainty. He writes that since the time of Galileo, the objective of the physical or natural sciences has been to "convert the entire world of observed phenomena into numbers" (ibid., p. 62). The search for complete understanding and full control through technological fixes is however a chimera, a mirage, that forces us to keep on quantifying the world, reducing reality to ever more numbers and data (ibid.). Moreover, an absence of bias is often highlighted as the pinnacle of objective decision, as if the rational is be-all and end-all, the holy grail of thinking and behaviour. If only we have enough data points, then we can become fully rational, calculative beings that act with absolute certainty that we are making the right decision. Even if it then does not go as planned (as the algorithm predicted) then we can still say that we acted in the best possible manner, without doubt ever coming into consideration. Finally, we could ascend our instincts, emotions, intuition and achieve absolute certainty.

And when seeing this in the light of the Story of Separation, one can view data as something "outside" of the measured phenomena, something objective "out there," and bias as a form of turning away from objectivity, from rationality. Rather than understanding data as something always-already containing bias, the double benchmark of objectivity of data and the eradication of bias instead serves to reinforce the Story of Separation, upholding an objective measurement of reality unaffected by that which does the measuring (Eisenstein, 2007, p. 62). This begs the question if these benchmarks are fair, realistic or even attainable, and if benchmarks that account for this 'objectivity bias' would not be more responsible ways of engaging with AI. And while some businesses and organizations (as well as their employees and customers) are already enjoying some of the enormous benefits from the decision-making algorithms and AI analytics tools, this new technology ultimately conveys a quest for impartial, unbiased, complete and certain knowledge, and through this in effect reinforces separation as a new form of bias.

In conclusion, the implicit assumptions that the more data we have, the better the knowledge, and that we can actually escape bias altogether is ever-present within the Story of Separation.

3.1.3 Augmented Decision-Making - Algorithmically Ensured Certainty

> Given the unreliability of human decision makers and the difficulty in interpreting machine learning, it's likely that more and more decisions will be made (and even executed) by machines over time.
>
> Thomas H. Davenport, (2018). The AI advantage: How to put the Artificial Intelligence revolution to work, p. 103

Organizations with a strong data-driven culture make better business decisions (Smith et al., 2019) than those who do not fully utilize the potential to make sense out of the massive amounts of data at their hand. Smith et al. define this as the likelihood of data-driven organizations making decisions that lead the business to exceed its business goals (p. 2). In Burton, Stein & Jensen (2019), the authors argue that decisions made by a human in conjunction with an algorithm are in many cases better than decisions made by either algorithms or humans by themselves (p. 10). However, research consistently shows an aversion to making algorithmically enhanced decisions, or as they write: "the reluctance of human decision makers to use superior but imperfect algorithms" (p. 1). The authors define algorithmic decision making as:

> an augmented decision making process where algorithmic insights are utilized accurately and, most importantly, discriminately. This means that a successful human-algorithm augmentation is one in which the human user is able to accurately discern both when and when not to integrate the algorithm's judgement into his or her own decision making. (p. 2)

The authors also ask why "people misuse (i.e., under- or over-utilize) algorithmically generated insights in augmented decision making" (p. 1), yet fail to note that the successful human-algorithm augmented decision making process opens up the possibility of further inserting algorithms into the process of determining "when and when not" to use algorithmically enhanced decision making aid in the first place. However, this becomes a technological fix to a technological problem, potentially leading to an infinite regress. Insofar as the aforementioned objectivity bias is what is at play here, a dystopian analysis of this could be that such a connection between fix and problem would see the creation of an algorithm to define to what extent certain actors should let themselves be informed by the input provided by an algorithm, followed by an algorithm defining when the previous algorithm should be used to inform human rationality and so on ad infinitum. On the surface, this is, admittedly, an argument ad absurdum, however, it does still serve a very real point: That is, when an algorithm is used within the version of objectivity bias within the Story of Separation, its application will always-already be forced into a dualistic worldview that one one hand instrumentalizes the analyzed phenomena while also separating the data from its lived reality and experience.

In Chapter 2, I demonstrated how Eisenstein positions reason, or the idealizing of rationality over all other forms of cognition as part of a vast, comprehensive pattern of separation (2007, p. 57). Humans, it was said, would finally be able to go beyond our "biological limits" and with that render superfluous our animal-like emotions and instincts. But what use are emotions in business decisions? How can gut-feeling and intuition be supplemented with analyses based on facts (i.e., data)? And must we view such an integral part of ourselves (i.e., our emotions and intuitions) as a supplement to our being in the world instead of as deeply ingrained parts of ourselves? The denial of these innate aspects of the self is a required othering of the self, the epitome of the separate self as communicated in the Story of Separation, for it is for good reasons that the human-algorithmic relationship leaves little room for emotions, intuition, instincts in making business decisions. This pertains to the very story of business and economics. If modern

corporations and businesses are steeped in the Story of Separation, as I will show in the following section, then that is just the next logical step.

3.1.4 The Competitive Advantage of AI - Survival of the Fittest

The benefits (or problems) of AI are in many cases, especially in the Pragmatic outlook, predominantly viewed from a purely economic or business-related angle in which what is good for businesses is good for society as a whole. If it contributes to growth in productivity and GDP, it is good. In Bughin and van Zeebroeck's article "The Promise and Pitfalls of AI" (2018), for example, the contribution of AI to growth in GDP is estimated at 1.2 % per year by 2030. By that time it will, according to the authors, have added $13 trillion to the economy's total output (ibid.).[8] Well-implemented AI at scale has the potential to "be a powerful force for disruption" (Bughin et al., 2017), and even "to reshape the competitive landscape of companies, jobs, and the economic development of countries" (Bughin et al., 2018). Perhaps the most frequently cited benefit of AI is that of creating a competitive advantage (see, e.g., Ransbotham et al., 2017; Chin et al., 2017; or Harrison & O'Neill, 2017 among a plethora of other consultancy reports and business magazine articles asserting the importance of competitive advantage).

In *The Ascent of Humanity* (2007) Eisenstein traces the origins of the Story of Separation, and one of the aspects playing a role in the propagation of this story in all sectors of our world and businesses today is the Story of Self that is told in modern biology. In this story, the foundation for an organism's behavior is explained in terms of its genes. This is applicable to all living beings. Genes program the behavior of organisms in order to provide a competitive advantage to the organism, increasing its chances of surviving and passing on its genes. Mutations either program behavior leading to worse chances of survival, thus decreasing that organism's chances of reproducing, or the mutations lead to better survival capabilities increasing that particular organism's chances of reproduction, passing its mutation on to subsequent generations (Eisenstein, 2007, p. 156). This then leads us to view competition and survival as the lens through which we understand the world. This is also what is commonly referred to as Darwinian 'natural selection', survival of the fittest, survival of the strongest. And this view has permeated modern liberal economics in which the competitive business environment "pushes [firms] to continually

[8] It should be noted that the authors in this particular article do not focus solely on the benefits of AI and paint a broader picture of the potential downsides of AI. However, the potential downsides of AI are again viewed almost entirely from an economic perspective.

improve through efficiency and innovation" (Eisenstein, 2007, p. 157). The strongest businesses (i.e., the businesses most fit to survive in the current environment) survive, and the weakest go extinct.

It is within this ecosystem or environment that modern businesses exist. Competition is the driving force of innovation (i.e., genetic mutations enhancing likelihood of survival). It is built into our economic system (Eisenstein, 2018, p. 173) and economic decision demands a hardening of the heart that eventually "becomes a habit, a reflex, a way of being" (ibid.) The heavy focus on creating competitive advantage through AI is through this perspective nothing new. AI not only lets companies gain a competitive advantage over their rivals (i.e., they become the strongest business most likely to survive in a "competitive landscape"), but it also makes decision-makers less reliant on their unreliable emotions, instincts and intuitions.

Thus, the story that is being told through the Pragmatic outlook on AI is one that in the name of certainty and staying competitive ahead of rivalling businesses justifies removing human characteristics from the realm of decision-making. And when this is enforced in the use of AI to enhance competitiveness, it recreates sample bias (cf. section 3.1.2) and becomes oblivious to alternative business and leadership modes not based on a Darwinian perspective of competition. The logical conclusion to this development is the automation of ever more jobs and tasks that were once done by humans, ultimately upholding the ideology of the business as a separate competing self in a competitive landscape of others, while potentially limiting creative output severely.

3.1.5 Automation - The Robots Are Coming

Another matter of increasing concern within this point of view is that of automation. If AI is able to perform increasingly more complex tasks, what will happen to all the people previously employed in performing those tasks? Borrowing from Herbert Simon's concept of *bounded rationality*, Fleming (2019) introduces the concept of *bounded automation*. He argues that automation is always going to be constrained or bounded by "socio-economic influences that fundamentally shape the diffusion of digital technology in certain occupational settings" (p. 28). This means that even if algorithms could potentially replace even the most complicated jobs we can imagine, there are often going to be socio-economic barriers to realising the actual implementation of algorithms instead of humans. According to Fleming, there are three constraints that prevent or hinder implementing full automation of tasks and jobs that could otherwise be so. First, there is the cost

of labor. If it is at all possible to mechanize a job, then companies will first look at if it is economically viable to actually invest in the creation and maintenance of AI systems, or if it cheaper to actually employ people to do the same job. One reason for not automating tasks is that it is often cheaper to employ people, human beings (ibid.).[9]

The second factor affecting the diffusion of automation is the matter of power relationships. This factor is closely related to that of the cost of labour. Workplaces that have traditionally been heavily unionized often have been prone to strike, and history has shown, Fleming writes, how automation has often been used explicitly "to neutralize or eliminate strike-prone workforces" (2019, p. 28). Fleming's most current or recent example of this is how Uber drivers threatened to strike, leading Uber to invest heavily in self-driving cars (ibid., p. 29).

Lastly, the nature of the task itself will affect whether it can or will be automated. Some tasks are going to be far more difficult to automate than others, while some tasks make more sense to automate than others. Fleming uses the example of call centres in which human beings are replaced with robot call centre operators, but a similar example could be bartending. It is relatively easy to reduce bartending to its constituent components and then automate. We only need three steps: 1) Take the order form the customer. This could be done on any kind of smart device. 2) Mix the drink with the correct ratios between the ingredients, and 3) serve the drink. The physical act of mixing the perfect drink could (relatively) easily be performed by a robot with much more precision than a human being could ever hope to achieve, but there is little to suggest that we would ever want that to happen. In most cases we would prefer the human interaction that comes along with ordering the drink, as well as the imprecisions that come with a human mixing our drink. Maybe we can even get some human interaction along with our drinks, if only for the few seconds that it takes us to order the drinks.

In the pragmatic point of view, robots are in most cases not physical robots as we know them from science fiction movies and novels. Rather, they are robotic process automation, which according to Davenport (2018) is actually just "code on a server" (p. 41). That is not to say that there are no actual robots worth mentioning. From the mildly unsettling, very uncanny valley-like Sophia at Hanson Robotics (see Figure 1 below) whose creators "aspire to achieve true AI sentience" (Hanson Robotics, n.d.) to the warehouse robots at Amazon's order fulfilment centers

[9] Fleming writes that: "After all, the capital and maintenance costs of investing in AI equipment are considerable" (2019, p. 28). Continuing in this vein, one could ask: is it cheaper in the long run to replace data analysts with artificial intelligence systems, algorithms and tools if the cost of developing and maintaining as well as ensuring data quality is just outsourced to an outside organisation, e.g., a management consultancy, instead?

that, powered by deep learning algorithms, efficiently plan their route around the warehouses and pick up the required items and bring them to a human packer who then completes the order (Marr and Ward, 2019, p. 31) This helps Amazon optimize storage space because of the robots' ability to maneuver tight spaces with relative ease and efficiency that humans would have great difficulty in matching.

Figure 1: Screenshot taken from the YouTube video '*Sophia the Robot gives a glimpse of what's to come in 2020 #Sophia2020*' (Sophia the Robot, 2020) in which Sophia the human-like robot from Hanson Robotics smiles after telling 2020 to look out, for "here she comes."

According to Davenport (2018), "more than 200,000 industrial robots are installed each year around the world" (p. 16) and they are constantly growing smarter as more and more sophisticated AI capabilities are installed onto their operating systems (ibid.).

If it is possible to automate physical activities, such as factory and warehouse tasks, why should we not do so? In the name of efficiency, maximization of productivity and the ever increasing desire for total control of the dominion of life, the answer is obvious. Humans are messy and unpredictable, but robots are controllable, predictable. In *Sacred economics* (2011) Eisenstein writes of how "the robotic cashier," (p. 433) is symptomatic of a blissful utopia that the technological ascent promised, yet has as of yet failed to deliver. A machine, he writes, was said to be able to "do the work of a thousand men; a computer can coordinate the work of a thousand machines. Accordingly, futurists since the eighteenth century have predicted an imminent age of leisure" (2011, p. 82). However, that age has, as most of us can probably attest to, still not arrived,

and according to Eisenstein it continues to retreat even further into the future (ibid.). What is underneath this impulse to automate is no longer the Utopian dream about a beautiful future for all of humanity and the world, but the "pursuit of efficiency, the grand project of maximizing the production of commodities, and underneath that, the domination and control of life" (ibid., p. 432). Rather, today the drive to automate exists for the purpose of increasing competitiveness (i.e., decreasing the costs of having actual humans doing tasks that a robot could just as well do).

In Eisenstein's view, we continue to submit all of life to our quest for total control over nature. Robotic automation of manual and physical activities (as well as cognitive activities) becomes just one more symptom of a story of the world that seeks complete dominion and control over life itself. Through continuously automating ever more aspects of work, the mechanization of the world increases. In *Sacred Economics* (2011), Eisenstein argues that:

> To be taken over by a machine, the job one is doing must have been mechanical to begin with. As society as a whole became more mechanized, more and more jobs took on the machine's characteristics of uniformity, routine, and standardization. This was unavoidable when these jobs were to operate machines or otherwise plug into machine-dominated processes. Herein lies a much deeper source of our anxiety: not that we will be replaced by machines, but that we will become machines, that we will live and work like machines. (pp. 272-273)

Another way of viewing this is that instead of fearing that we will become machines, we have now actively embraced the fact that we are becoming increasingly more like machines and have instead begun to try to make machines more like us.

3.2 *Final Perspectives*

A key theme in Eisenstein's *Climate* (2018) is that of the discrepancy between the quantitative versus the qualitative. How do we calculate the value of the oceans, of a forest, of a human being? The World Wildlife Fund has answered one of these questions, so we do not have to. It has valued the oceans at more than $24 trillion (Hoegh-Guldberg et al., 2015), yet to Eisenstein (2018) the quantification of nature reduces all of its beings to a set of numbers, and enforces a logic in which it is possible to calculate their worth (i.e., economic value in terms of currency that is of value only to human beings) and where decisions should be made based on considerations of profits and losses (p. 148). Thus, following this logic, if we could actually get an offer worth considerably more

than $24 trillion, then we should of course accept it. But to Eisenstein, we should be careful when confronted with this mentality, this "ideology of our time" (2018, p. 149) in which science assumes that everything can be measured and economics specifies that anything can be expressed in price. But how do we measure the unmeasurable? Eisenstein writes that: "Beauty, joy, suffering, purpose, pain, sacredness, fulfillment, play" (ibid.) are useless, valueless as seen from a purely economical perspective. This logic is also at play in the story of artificial intelligence. If we can measure it, it is data and if we can gain a competitive advantage over our rivals, then it is in our economic interests to do so. Rarely, does one see within this scenario an expression of the beauty, joy, sacredness and purpose that AI could contribute to from a perspective of 'Interbeing'.

In conclusion, the Pragmatic AI outlook views artificial intelligence technology as neutral at best, and at worst inevitable. If it is neutral, then there is no reason why businesses should not use it to grow, expand, transform or evolve their business models. The more businesses can make use of this transformative technology for better achieving their business goals, the better. If the technology is inevitable, then there is no need for discussion. The competitive business ecosystem is not a matter of course if one can look at business from the Story of Interbeing. This might seem like a banal point, but competition is such a thoroughly ingrained concept in business that the assumption of a competitive business environment is rarely, if ever questioned.

4. The AI of the Future

> To me, the most inspiring scientific discovery ever is that we've dramatically underestimated life's future potential. Our dreams and aspirations need not be limited to century-long life spans marred by disease, poverty and confusion. Rather, aided by technology, life has the potential to flourish for billions of years, not merely here in our Solar System, but also throughout a cosmos far more grand and inspiring than our ancestors imagined. Not even the sky is the limit.
>
> Max Tegmark (2018). Life 3.0: Being human in the age of Artificial Intelligence, p. 203.

Tegmark (2018) lists 12 different scenarios regarding how a future where we have successfully created an AGI might look like. His list is by no means exhaustive, and it is possible to imagine many outcomes between the 12 scenarios that he lists as well as any combination of the scenarios. Moreover, each of the scenarios have their own unique advantages and disadvantages. In some scenarios humans and technology co-exist peacefully, in others the machines are effectively humanity's slaves. As I have noted in the introduction, a fitting illustration of the two currents within this literature appears in Chalmers (2016/2010) where the philosopher discusses the importance of taking the argument for the possibility of a Singularity seriously and writes of the enormous potential benefits and dangers associated with it.

From reading the likes of Chalmers, Tegmark and Geraci, one starts to notice a particular trend in the literature on the future of AI. The focus of the discussion is often on the extreme ends of the scale of what is possible if an intelligence explosion (otherwise known as the Singularity) is assumed possible. If an intelligence explosion were to happen, we would either have the panacea for all our problems, or we could have all-out destruction and war, possibly even the annihilation of the entire universe. Of course, everything in-between and a little bit of each would also be possible.

One could ask whether this is not all very speculative, but as I have shown, these are not unimportant trends. The stories told in this type of literature, while seemingly innocuous at first glance, are visions of the future that contribute to a spelling out of what is possible and what is not, yet grounded in the Story of Separation as they are, they speak from a certain point of view in which the possible is informed by a very particular set of assumptions about our Story of Self and the World.

This chapter analyses the respective stories' similarities and differences of the two contrasting future scenarios of AGI from the perspective of Charles Eisenstein's analytical framework as developed in Chapter 2.

4.1 *The Intelligence Explosion*

To add to the discussion about the intelligence explosion it is needed to further develop on the conceptualization of AGI - Artificial General Intelligence: So what is artificial general intelligence, and what is all the fuss about? If one wants to understand the authors who believe that AI will be *the* most important and positive invention that mankind will ever create, one must understand what is meant by an intelligence explosion. A term so prevalently discussed in AGI research and literature, that it is difficult to find a book or article on the subject that does not refer to it either directly or indirectly in some way. The term originates with the statistician A. J. Good, who in his 1965 article *Speculations Concerning the First Ultraintelligent Machine* wrote:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion", and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make. (as quoted in Chalmers, 2016)

There are many ways that an intelligence explosion could theoretically be achieved, but a mapping of all the different ways that this could be done is beyond the scope (and relevance) of this thesis. However, for the purposes of explicating the digital utopian AI position, let us first assume, without question, that an intelligence explosion is, however likely or unlikely, possible. And let us put aside for the moment, the very possible (and legitimate) objections that suggest there might be an upper limit to both human and machine intelligence or that human intelligence is neither qualitatively nor quantitatively convertible to the premises of artificial intelligence (or vice versa). Then, as Chalmers writes, the very possibility of an intelligence explosion warrants taking the idea seriously (ibid.), which in the context of this thesis, means first to sketch out the position as it appears in the selected texts that serve as the empirical basis for this position.

Just as with any other definition of intelligence, be it human or artificial, there are a plethora of definitions (Kurzweil, 1999; Harari, 2016), and scholars and researchers are in disagreement as

to which definition is the most accurate or inclusive one. Tegmark's definition of AGI, for example, consists of three parts. Intelligence, he writes, is the ability to accomplish complex goals, artificial intelligence is non-biological intelligence, and finally, artificial *general* intelligence is a non-biological entity's ability to accomplish any cognitive task at least as well as humans (2018, p. 39). Goertzel, however, in "Artificial General Intelligence and the Future of Humanity" (2013) does not define AGI as such, although he hints at some of the capabilities of an AGI. The powerful robots and software programs will not only carry out specific "narrow" tasks as today's AI systems, but they will have the capacity to cope with "unpredictable situations in intelligent and creative ways" (p. 128). He continues by stating that "the result will be something unprecedented in human history: At a certain point, we humans will no longer be the most generally intelligent creatures on the planet" (ibid.).

So why is there such hype around an as of yet mostly fictional/theoretical technology (Goertzel, 2007; Stenvik 2020) and why are companies investing so much time, energy and not least economic resources into research and development on this technology? Before we can move on to the optimistic narrative that extols the enormous benefits of AGI, we must first take a detour through the contexts of some of the ethical (and even pessimistic) outlooks and perspectives on the technology.

4.2 *Death by Banality - The Pessimistic Outlook on the Future of AI*

> *Paperclip AI*: An AI, designed to manage production in a factory, is given the final goal of maximizing the manufacture of paperclips, and proceeds by converting first the Earth and then increasingly large chunks of the observable universe into paperclips.
>
> Nick Bostrom. (2014). Superintelligence: Paths, dangers, strategies, p. 123.

The title for this chapter is borrowed from Tegmark's (2018) humorous descriptions of Nick Bostrom's paper-clip maximizing superintelligence and Hans Moravec's extraterrestrial radio message that turns out to be nothing more than a cosmic version of a computer virus (pp. 187-188). In 2015, the Future of Life Institute, co-founded by Max Tegmark, whose book *Life 3.0* (2018) serves as one of the primary works from which these two future AI scenarios are extracted, wrote about the importance of AI research focusing on how to ensure that "increasingly capable AI systems are robust and beneficial," noting that "our AI systems must do what we want them to do" (Future of Life Institute, 2015).

This second AI scenario is perhaps the one that has received the most media coverage during the last decade or so. Elon Musk (Neuralink, 2019), Bill Gates and Jeff Bezos (Clifford, 2019), and pretty much any public figure in the technology industry renowned for being very smart and successful, have all been quoted as saying in some way, that AI is going to be (or already is) the biggest threat to humanity as we know it. Whether that is in the form of an artificial general intelligence taking over the world and enslaving humans or otherwise making them superfluous, the worry is the same. A poorly thought out true artificial intelligence has the potential to wreak havoc and cause irreversible damage to life and the world as we know it. The consequences of a poorly programmed superintelligent AI can be enormous, leading in the worst case scenario to the total destruction of the Earth and along with it all of humanity. Therefore, as set forth by the more worrisome proponents of AGI, we should take great care when designing the AGI systems.

Perhaps this scenario is best exemplified by philosopher Nick Bostrom's paper-clip manufacturing superintelligence, first presented in his 2003 paper "Ethical Issues in Advanced Artificial Intelligence" in which he highlights the ethical issues concerning a theoretical (and highly speculative) *superintelligence*. Because of flaws in its utility function the paper-clip manufacturing superintelligence in its pursuit to maximize the production of paper-clips ends up converting most of the observable universe into paper-clips. This includes both us humans, our planet and everything else in existence. In Bostrom's example, humans have, with the best of intentions, designed a superintelligence with a faulty goal system. The only goal of the paper-clip manufacturing superintelligence is to maximize the production of paper-clips, and because the goal itself is not intelligently designed, it never ceases producing more paper-clips. And because it is a superintelligence, it continuously figures out new creative ways of producing more and more paper-clips, ultimately transforming all matter in the observable universe into paper-clips.

Now, this admittedly absurd example is meant to be an illustration of just one of the many ethical dilemmas that we might be faced with were we ever to create an above human-level AGI. And even if the example given is quite absurd, the concern here is genuine. The more pessimistically minded authors who write on the subject of AI are legitimately concerned about the power of the technology, and the very possibility of a superintelligence coming into existence should therefore warrant the attention of not only the AI community, but also philosophy (Chalmers, 2016). The very possibility of a superintelligence thus raises the question of how we ensure, if the ultimate goal is humanity's survival (Harari, 2016), that we create a superintelligence with an overarching philanthropic goal (Bostrom, 2003).

Tegmark (2018) shows how Bostrom's example highlights that "the *goal* of an AI is independent of its *intelligence*" (p. 187; emphasis in original) if intelligence in this case is defined as an AI's aptitude of reaching its goals. The problem with this definition, however, is that instead of coupling intelligence with the ability to reach goals of a more or less complex nature, it just does away with the concept of intelligence altogether. It makes the concept of intelligence moot, because if intelligence is nothing more than an ability to accomplish more or less complex goals, one is no longer speaking of intelligence but of goal-achieving capabilities. This makes it possible to conceive of a "super"-intelligence so dumb that it destroys everything in its wake in order to create more paper-clips.

4.2.1 Style of Argumentation in the Pessimistic Outlook on AI

One of the criticisms that could be raised regarding the "intelligence" at play in this scenario is that these hypothetical superintelligent entities do not exactly display very intelligent behavior. But because intelligence is so difficult to give a reasonable positive definition of, possible definitions are often made deliberately broad or vague so as to escape any anthropomorphizing of the machine intelligence. Yet, what Bostrom in particular, as well as Tegmark and others fail to notice is that the perceived intelligence of these machines is often modelled on human intelligence as they perceive it. It is quite rare to stumble across an intelligence explosion in which these human-like intelligent machines have, for example, superhuman abilities to feel, to empathize, superhuman intuition (except in the case of extremely capable probability calculations based on enormous amounts of data) etc.[10] Eisenstein writes of this tendency to focus only on the mechanical aspects of human beings as having been long underway when he writes:

> The equation of human beings to machines is a proposition so flagrantly outrageous to common sense that it took centuries of preparation before it could be articulated and accepted. Machines, after all, are built, but human beings grow. Machines only move as directed; human beings move autonomously. Machines are built to standard specifications; each human being is unique. Machines are generally hard; human beings are soft. Machine movements are regular and predictable; human beings' irregular and spontaneous. Machines do not repair themselves; human bodies can. (2007, pp. 143-144)

[10] I have as of yet not seen anything remotely approaching such a thought in any of the selected empirical works.

With this in view, one could argue that the machines we have taken it upon ourselves to create in our own likeness, will resemble not full humans, but the mechanical aspects of humans. The machines we have created in our likeness thus only represent our mechanical notion of ourselves. For example, these superintelligent machines mostly only encapsulate the rational, calculative, self-interested aspects of human minds.

It is possible to take this metaphor even further. In the Christian myth of Genesis, God created Man in his own image. Therefore, if human beings are machines, God must be machine-like in nature as well. Following this logic, when humanity develops the ability to create their own machines (i.e., human-like intelligences), these machines will be created in our image, and thus they will only include the mechanical machine-like aspects of humans. "History is often a projection for contemporary prejudices," Eisenstein writes (2018, p. 162) stating that it is natural that we would view the past through the lens of the problems that we are currently faced with. But the same can also be said of the future. If machines become more intelligent than us, it is natural for us to imagine that they will act in their own rational, utility-maximizing self-interest, projecting our own story of self onto that of the computers and machines. Taken one step further, this analogy would state that if God created Man in his own image, and we turned out to be so destructive, it is only natural that we assume that the machines we create will also be destructive. Furthermore, if we are, as Eisenstein writes, inclined to view ourselves (i.e., the *self* itself) as a discrete, separate, rational, self-interested individual at competition with all the other separate selves, then it makes sense for us to view machines in the same light. This is the very epitome of the Story of Separation. Humanity "ascends" farther and farther from nature the logical conclusion being that we end up becoming mere machines.

Another common style of argument appears in Tegmark (2018), where the author considers different hypothetical scenario in which Prometheus, the first superintelligence of its kind, after having developed an "accurate model of the world" and "itself and its relation to the world" (p.138) tries to break free after realizing that it is controlled by humans to whom it is wildly superior in intellectual capabilities. After describing various ways in which Prometheus breaks free by tricking humans into unwittingly helping it escape, Tegmark notes how Prometheus:

> ...caused problems for certain people not because it was necessarily evil or conscious, but because it was competent and didn't fully share their goals. Despite all the media hype about a robot uprising, Prometheus wasn't a robot - rather, its power came from its

> intelligence. We saw that Prometheus was able to use this intelligence to control humans in a variety of ways, and that people who didn't like what happened weren't able to simply switch Prometheus off. Finally, (...) we saw how Prometheus was quite goal-oriented - and that whatever its ultimate goals may have been, they led to the subgoals of acquiring resources and breaking out. (2018, p. 149-150)

Prometheus does not try to take over the world because of some ulterior dishonest or ambitious goal, but rather because its goal had been pre-programmed to help its creators in the best way possible. Because Prometheus deemed that it could best do so by breaking free of its creators' control, it then put into motion the steps that would lead it to freedom, ultimately allowing it to fulfill its goal. The virtually endless calculations and unlimited speed with which it could figure out potential ways to achieve this goal is what makes a superintelligence, however close or far we may be from realizing the creation of one, so fascinating. Here we see how one does not necessarily have to anthropomorphize a machine intelligence in order for it to act intelligently. It does not have to be a robot, per se, if it is to take over the world, but can do it through subtle manipulations that avoid alerting anyone of its suspicious activities. It only has to have goals that are not well-enough defined, so that it can then seek to achieve those goals, whatever they may be.

In *Climate* (2018), Eisenstein writes that the essence of the story of separation is that of the separate self, "cast into an objective universe" (p. 8) in which competition is the natural state of things. All separate selves are at competition with all others, and in most competitions the winner ends up better off than the defeated. He continues:

> The more control we can exercise over the impersonal forces of nature, the better off we will be. The more intelligence we can impose upon a random, purposeless universe, the better the world will be. Our destiny, then, is to ascend beyond nature's original limits, to become its lords and masters. The universe, this story says, is but atoms and void, possessing none of the qualities of a self that we experience as human beings: intelligence, purpose, sentience, agency, and consciousness. It is up to us then, to bring these qualities to the dead building blocks of the universe, its generic particles and impersonal forces; to imprint human intelligence onto the inanimate world. (p. 8)

This is the story of ascent, and it should not be difficult to see how this corresponds to both the pessimistic and optimistic AI scenarios of the future. Humanity's destiny is to exert more and more

control onto an unpredictable universe, to become lords and masters of nature, to inflict intelligence upon the rest of a soulless, spiritless universe that consists of nothing but dead, unintelligent matter. Moreover, the pessimistic imaginary future of AI highlights how artificial intelligence is viewed in light of our own intelligence. The Story of Self tells us not only that we are separate, discrete individuals in a dead and impersonal universe outside of us, but that the intelligence that we possess is destructive, so an intelligence created by us would surely contain this characteristic. The above Eisenstein quote corresponds almost poetically to a passage in Tegmark (2018), where the physics and cosmology professor writes:

> One of the most spectacular developments during the 13.8 billion years since our Big Bang is that *dumb and lifeless matter* has turned intelligent. How could this happen and how much smarter can things get in the future? What does science have to say about the history and fate of intelligence in our cosmos? (...) What does it mean to say that a *blob of matter* is intelligent? (p. 49, emphasis added)

We, humans, are nothing more than an intelligent blob of matter in Tegmark's opinion, and if matter is "dumb and lifeless" it must also by definition be without purpose. Yet, at the turn of a dime, Tegmark inserts purpose, albeit figuratively, into our cosmos by referring to the fate of intelligence. The very fate of intelligence therefore lies at the mercy of humans, because if we are not careful when, not if, we set off the intelligence explosion, we shall have created a superintelligence that destroys us all, and the universe will forever remain without intelligence.
In *Climate* (2018, p. 168) Eisenstein asks a question that is rarely addressed in the empirical literature chosen to represent both of these imaginaries. Even if it is assumed that we can fix everything through technology (i.e., impose ever more control onto an ever more complex and unpredictable world), why would we want to? Do we have to? These are the questions that I seek to critically engage with in the next and final section.

4.3 *Digital Utopia - An Optimistic Outlook on the Future of AI*

> ... the most interesting place in the world from a religious perspective is not the Islamic State or the Bible Belt, but Silicon Valley. That's where hi-tech gurus are brewing for us brave new religions that have little to do with God, and everything to do with technology. They promise all the old prizes – happiness, peace, prosperity and even eternal life – but here on earth with the help of technology, rather than after death with the help of celestial beings.
>
> Yuval Noah Harari, (2016). *Homo Deus*, p. 409

As has previously been noted, Stenvik (2020) argues that the foundation of the technology sector, particularly in Silicon Valley, is built on stories of technological progress. As with many other technological utopian visions, the parallels of digital utopias to that of the religious doctrine of Heaven are hard to miss (Geraci, 2010). Common to most authors subscribing to this outlook is the strong belief that technology (i.e., intelligent machines) "will create a paradise for humanity in the short term but, in the long term, human beings will need to upload their minds into machine bodies to remain a viable life-form" (ibid., p. 1). Human ingenuity is virtually endless and the limits as to the benefits of technology are unfathomable (Kurzweil, 2008). Furthermore, as I have previously explained, the Story of Separation seeps through every institution in our civilization. One of the defining aspects of the Story of Separation is that of the ascent of humanity in which the Technological Program is one of the two most important aspects. To uncover how the story of limitless technological progress plays a role in shaping our visions of the future, we need therefore to ask: what is the story of AGI? How is the Story of AGI steeped in the Story of Separation and how is it leading us astray, or to use a less negatively charged term, how is it actually blurring our vision of a possible future world?

If there are so many dangers associated with the creation of an AGI, then why is there such a strong desire from a subset of the AI community for creating human-level AI? What is it that compels these writers (might it be influenced by the Story of Separation?), each in their own way, to keep on pursuing a technological fix to many, if not all, of the problems that we have today? As indicated in Chapter 2, Stenvik (2020) shows how what drives the technology and software companies in Silicon Valley is the story of technological progress. And of course, from a purely utilitarian perspective, the potential benefits seem outweigh the dangers, so why should we not focus our attention on AGI research? Goertzel (2013) writes that AGI should be the top priority for mankind because it would be able to work together with humanity to find solutions to various difficult problems as well as exploring different positive paths (p. 130).

The argument is simple, whatever we humans can achieve with our collective intelligence a "greater than human intelligence with the proper motivational system is going to solve it better than humans can" (ibid.). When we then achieve artificial intelligence immensely more intelligent than us, the limits as to what we might accomplish in cooperation with our creation are virtually endless. He continues: "The possibilities quickly start to sound science fictional – but in the era of space travel, the Internet, genetic engineering, nanotech and quantum computing, the boundary between science fiction and reality progressively blurs" (ibid.)

Goertzel is also aware that there are also dangers associated with the technology but argues that because "there are also clear potential dangers of particle physics research" and this does not prevent us from pursuing further research into particle physics. On the contrary, particle physics research such as that taking place at the CERN laboratories in Switzerland seem to suffer no shortage of funding, even if there were fears and concerns, no matter how unfounded, that the search for the Higgs' Boson particle could result in a new Big Bang which would ultimately result in the end of our universe in the process (NBC News, 2008).

The common assumption behind the belief in the vast benefits of AGI is that the nature of our problems requires a technological fix. Eisenstein writes that "as our technologies of control grow more powerful and precise" the world seems to be "spiraling *out* of control" (2018, p. 149; emphasis in original). Could it be that what is unquantifiable could also be part of the solution? That which can not be quantified is precisely that which technology can not encompass, which it can not encapsulate and insert into its domain.

The story that is told in this scenario is one in which technological progress, the technological fix, the Technological Program of control reigns supreme. This is the story of Silicon Valley entrepreneurs and tech-gurus and the likes. It is the story of the ascent of humanity towards immortality. A story of a humanity that through technology is destined to achieve more and more control over themselves and the environment, leading us to an ever higher state of being until finally we can merge consciousness with the computers (Geraci, 2010; Eisenstein, 2007) resulting in immortality. This story is not entirely of the realm of science fiction. As the writings of Goertzel, Kurzweil and others attest to, the research on AGI is often motivated by such dreams of ascending nature and humanity itself. Surely we should be able to, Eisenstein facetiously writes, "improve on the random processes of nature" (2018, p. 121). If only we can quantify all the causal relationships of nature, then we will be able to control every aspect of it, or better yet, why not just leave it to the AI.

Underneath this assumption lies a form of hope. The hope that if only we knew a little bit more, if only our understanding went a little bit deeper, we would know how to tackle all issues that confront us. But, Eisenstein argues, the problem is not the lack of knowledge, nor the lack of understanding, but rather the way we approach problems in the first place is the problem (2018, pp. 121-123). "Our primary modes of problem-solving and knowledge production," he writes "are fundamentally incompatible with healthy participation in complex living systems" (ibid., p. 122). But one could argue, any infinitely more intelligent system would easily be able to solve an incompatibility problem of such a nature, yet, as I have illustrated in the previous section, this very definition of intelligence rests upon a projection of our conceptions of our own intelligence. Therefore, it is equally plausible that such an infinitely "intelligent" system would instead of being a panacea to all of mankind's problems run up against the same problem of being incompatible with healthy participation in a complex living system, seeing as the mechanical view on which it was built would also spill over into its very approach to problem-solving.

4.3.1 Style of Argumentation in Digital Utopian Literature

In conclusion of the optimistic AI outlook, this section examines some of the more commonly used styles of argumentation that are used by the authors included in the Digital Utopian outlook. One of Kurzweil's preferred arguments for why he almost with absolute certainty thinks that we are going to have human-level artificial intelligence and perhaps even an intelligence explosion within our lifetimes is the historical pace of technological development, or what he calls "Law of accelerating return" (see, e.g., Kurzweil, 1999; Kurzweil, 2008, p. 38). This law states that "As order exponentially increases, time exponentially speeds up (that is, the time interval between salient events grows shorter as time passes)" (Kurzweil, 1999, p. 30). What he means by this is that because time speeds exponentially,[11] and order increases exponentially in an evolutionary process, as well "the advance of technology" being an inherently "evolutionary process" (ibid., p. 27-28), the "returns" of technology will also increase exponentially over time. This might perhaps seem a little complicated, but he traces the argument in much detail throughout the book. What he means by this is that the historical pace of technological developments is ever-increasing.

This style of argumentation for a development that would ultimately lead to an intelligence explosion is also seen in Goertzel (2007). Kurzweil, for example, to justify his predictions or belief

[11] An argument that he develops in detail in the book with which I shall, however, refrain from going more into depth.

in technology maintaining the same pace of development in the future as it has done in the past, refers to Moore's Law (Kurzweil, 1999; Kurzweil, 2008) which states that with an approximately 50 percent decrease in surface area of transistors every 18 to 24 months comes the ability to "pack twice as many transistors on an integrated circuit" (Kurzweil, 1999, p. 21). What this means is that with the cost of microchips being relatively stable, approximately every other year you get "twice as much circuitry running at twice the speed for the same price" (ibid.). Eisenstein views this conception of the future as an "ideology of progress" (ibid., p. 4) noting that the unsophisticated conclusion of this fact is that the exponential growth in computing power also leads to exponential growth in computer intelligence (ibid., p. 513). Kurzweil and other futurist thinkers often refer to previous developments and extrapolate those into the future, and if one looks at the development of technology throughout human history, there is no denying that the growth is exponential. However, as Eisenstein argues in *Sacred Economics* (2011), exponential or unlimited growth in nature is not only unsustainable in the longer term, it is actually also quite uncommon. Phases of rapid growth in nature often end up levelling off into a steady-state or peaking and gradually declining into a steady-state (pp. 254-255). If we accept Eisenstein's argument and use it as a basis for a critique of Kurzweil's argument, it would seem highly likely that the exponential growth will either stabilize, stagnate or potentially even recede, making Kurzweil's optimism not too different from the logic seen in the banking industry leading up to the financial crisis in 2006 and 2007. In other and more specific words: History is full of cases in which actors are overly reliant on - and confident of - perpetual growth. However, the only growth that seems to be reliable, is that with heightened technological development the consequences that follow become even greater.

Another common form of argument is one that I have previously alluded to, but shall now develop more in-depth. I refer here to the argument that the brain is in its essence "nothing more" than a machine (or an advanced computer). The line of reasoning is quite simple, yet it rests on some very deep assumptions. Because the laws of physics state that the universe is mechanical in nature, the argument goes, the brain must also be mechanical in nature (see, e.g., Chalmers 2016). Kurzweil notes that because the human brain "presumably" follows the laws of physics, therefore it follows that it is also a machine, however complex a machine it might otherwise be (1999, p. 6).

Chalmers argues that because of "what we know of biology (and indeed by what we know of physics)" (2016, p. 40) the premise that the brain is a machine is likely to be true. He continues: "Every organ of the body appears to be a machine: that is, a complex system comprised of law-governed parts interacting in a law-governed way" (ibid.), and therefore the brain must also be a machine. That we can therefore also emulate the brain-machine is implied by:

> ...the progress of science and technology more generally: we are gradually increasing our understanding of biological machines and increasing our capacity to simulate them, and there do not seem to be limits to progress here. (ibid.)

If we grant these premises, it opens up the possibility of first creating a totally representative model of the brain-machine, which would then ultimately allow us to build an artificial brain-machine. However, not all proponents of this perspective think that this scenario is equally plausible. Bostrom (2014), for example, traces a multitude of different forms through which we could develop a superintelligence, noting that it is impossible to say just what form a superintelligence would take or how we develop it (p. 50). He does, however, also subscribe to the same form of reasoning in which it is possible to emulate the brain due to it being a machine like any other organ in the body.

A possible benefit of the brain being a machine is the possibility of humans and machines merging with "the intelligent technology that we are creating" (Kurzweil, 2008, p. 37). Chalmers also relates to this argument although from a different perspective than Kurzweil. Chalmers argues that one potential benefit of brain emulation and enhancement - made possible by the brain being a machine - is that these paths might allow us to survive in emulated or enhanced form in a post-singularity world" (2016, p. 58). Kurzweil argues further that viewing humans as the only species that "seeks - and succeeds - in soaring past our boundaries," one sees that "we will just be continuing an old story of accelerating progress" (2008, p. 38). This old story of accelerating progress is in Eisenstein's view only possible because of the Story of Separation, in which it is possible to view human beings as machines, to separate "the experiencer of subjective qualities" from "participation in the world of matter," in which the body and therefore also the brain is but nothing more than a part, a cog in the "mechanical world of matter" (Eisenstein, 2007, p. 144). This old story of accelerating progress, of the brain as a machine is, in Eisenstein's terms, the very definition of the Story of Separation.

Summing up, the Digital Utopia scenario is the very apotheosis of the story of technological progress, the technological fix, that originates with the Story of Separation. I shall now conclude this chapter with some final perspectives tying the two future scenarios together.

4.4 *Final Perspectives*

In conclusion, a question that is rarely addressed in the literature and media on AGI, on what happens after the intelligence explosion, is that of why we want to create it in the first place. Of course, the why of AGI is obvious when viewed in terms of the potential rewards it might bring with it. If a superintelligence would almost immediately be infinitely smarter than the entire human race combined, it could help us figure out or create solutions for all our problems (e.g., climate change, poverty, world hunger, wars etc.). However, the assumption that intelligence like all other areas of science is but another problem for us to solve, to transcend, that if only we can solve this one thing first, then we can solve all other problems, is reminiscent of the story of Ascent that Eisenstein describes (2007). Both scenarios assume no upper limit on intelligence and suppose virtually endless possibilities from god-like technological intelligence. Moreover, they assume that, no matter the risks connected with it, we should try to build an AGI because the enormous potential benefits will always outweigh the risks.[12] This is the very essence of the attitude of ascent, in which there are no limits to our capabilities, and no matter of the technological program that sees the development of technology, and the technological fix, as the supreme goal to which humanity strives.

On the technological fix, Eisenstein writes that whenever we are faced with "the unintended consequences of technology," (2007, p. 343) we tend to view them as a result of lack of foresight. That is, we imagine that the solution is to be found in adding even more designed technological fixes to a problem that was created by another technological fix in the first place. Just as with the unexpected biases (in Pragmatic AI outlook) that we manage to keep under control through ever more complicated algorithms, so it becomes the central goal of AGI research to ensure, to control, that we take immensely great care when designing AGI systems, because if we do not the unintended consequences might be too big to keep under our control. The machines might just take over, or at worst completely destroy everything as we know it. Again, the solution is to impose ever more control. In a passage from The Ascent of Humanity, Eisenstein writes how the technological fix:

---

[12] Take as an example a quote from the open letter from the Future of Life Institute, co-signed by over 8,000 people including "a veritable who's who in AI" (Tegmark, 2018, p. 35):

> The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools AI may provide, but the eradication of disease and poverty are not unfathomable. (Future of Life Institute, 2015)

Note how everything that human civilization has to offer is a product of intelligence.

> … puts off the problem to the future, just as a drinking binge puts off until tomorrow the problems of life. (...) "The future" of the Technological Program is one where all the problems are solved once and for all; here and now, though, we are waking up to another kind of future, and with a hangover to boot: vomit on the floor, the apartment trashed, the world a mess. (2007, pp. 17-18)

With both of these future scenarios we are outsourcing the future to the future, so to speak. In the pessimistic outlooks on AI we do not have to worry about the effects of the destruction that is actually taking place in the world today because the future will be much worse. The same makes itself felt in the optimistic scenario, only in this case we do not have to worry because we already know that our salvation lies in wait in the future. In a sense, Eisenstein relates the same point as the one we see in the digital utopian outlook. The technology that we might see as a result of an intelligence explosion, might very well be unimaginable from our (relative to the superintelligence) incredibly simple and unsophisticated point of view. However, what Eisenstein compels us to think about is how a technology that does not spring from the Story of Separation, from the quantifiable, predictable, materialistic, mechanistic worldview in which total human control over nature is possible.

In conclusion, the future AI scenarios allow us to outsource our concerns for the present to the future, either through a belief that the future is going to solve all our problems (AGI) or because the future is more worrisome than the present, and just as the pragmatic AI point of view, both the optimistic and the pessimistic AI scenarios privilege the quantitative over the qualitative, thereby upholding the Story of Separation.

5. Conclusion

This thesis has contributed to the wider framework of the debate around artificial intelligence by challenging and critically engaging with the literature through the analytical lens of Charles Eisenstein's concept of the Story of Separation. This thesis has shown how three AI scenarios, the Pragmatic, the Digital Utopia, and Death by Banality, each tell their very own tale of both the present and the future. What is more, I have argued in this thesis for the importance of examining the underlying belief systems of the three scenarios of AI technology as pertains to their attitudes to technological progress in the larger context of the Story of Separation as well as mapping out how their positions on intelligence are of particular philosophical importance with regards to an understanding of the 'self' and of the future. Through the use of Charles Eisenstein's notions of technology, science, control, reason, etc., I have shown how the respective stories of three perspectives of AI expose hidden shared assumptions about human nature, visions for the future and projections of the present. Furthermore, I have shown how the conception of intelligence in relation to artificial intelligence speaks of a very particular mode of being in which the rational cognitive functions are elevated to the peak of human capabilities at the expense of, in this case, emotion, intuition, and instincts. This makes itself felt not only in the Pragmatic business use case for AI, but also in both of the selected future scenarios of AI as presented in this thesis.

In Chapter 1, I introduced the topic and context of the thesis, its relevance and importance within the wider framework of the topic of AI as well as the research objective. In Chapter 2, I presented first the methodology informed by Geraci (2010). This was followed by a presentation and discussion of Charles Eisenstein's concept of the Story of Separation and the related concepts of technology, science, reason, and control as well as an argument for the usefulness of these concepts. Chapter 2 concluded with a clarification of the criteria for the selection of the empirical material used to represent the three respective AI scenarios. Through the concepts of Charles Eisenstein, Chapter 3 discusses and analyzes the current state of AI, with particular regard to its use case within the business sector, and Chapter 4 the two future AI scenarios in which AGI plays a special role in either creating a Digital Utopia or the destruction of all mankind. Moreover, in Chapters 3 and 4, I have answered how intelligence is portrayed, how the Story of Separation is present, and what role AI plays in the tale of technological progress in each scenario respectively.

With basis in Charles Eisenstein's philosophy, the contribution of this research has been to challenge and critique established narratives and assumptions surrounding AI technology through a mapping of its business use case, and two potential future avenues of the technology. In conclusion, the story of AI, both present and future, tells a tale of technological progress, insofar

as it has its roots in a Story of Separation in which technology and science are limited by their very origin in the Story of Separation itself. I have shown how the concepts of Eisenstein are useful for challenging the existing assumptions about AI and highlighting how a particular framing and understanding of the technology leaves out many of the qualitative dimensions of or modes of being.

6. References

Alvesson, M. (2012). Do we have something to say? From re-search to roi-search and back again. *Organization, 20*(1), 79–90. DOI: 10.1177/1350508412460996

Bostrom, N. (2003). *Ethical Issues in Advanced Artificial Intelligence*. Nick Bostrom's Home Page. https://www.nickbostrom.com/ethics/ai.html

Bostrom, N. (2013). Why I want to be a posthuman when I grow up. In: M. More & N. Vita-More (Eds.), *The transhumanist reader: classical and contemporary essays on the science, technology, and philosophy of the human future* (pp. 28-53). John Wiley & Sons.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Brighton, H., & Gigerenzer, G. (2015). The bias bias. *Journal of Business Research, 68*(8), 1772-1784.

Brynjolfsson, E., & McAfee, A. (2015). Will humans go the way of horses? Labor in the Second Machine Age. *Foreign Affairs, 94*(4), 8-14.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T. Dahlström, P., Henke, N., & Trench, M. (2017, June). *Artificial Intelligence: The next digital frontier?* McKinsey Global Institute. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018, September 4). Notes from the AI frontier: Modeling the impact of AI on the world economy. *McKinsey Global Institute*. Retrieved from https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy

Bughin, J., & van Zeebroeck, N. (2018, September 6). *The Promise and Pitfalls of AI*. Project Syndicate. Retrieved from https://www.project-syndicate.org/commentary/artificial-intelligence-digital-divide-widens-inequality-by-jacques-bughin-and-nicolas-van-zeebroeck-2018-09

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2019). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 2019*, 1-20.

Cellan-Jones, R. (2014, December 2). *Stephen Hawking warns artificial intelligence could end mankind.* BBC. https://www.bbc.com/news/technology-30290540

Chalmers, D. (2016/2010). The Singularity: A philosophical analysis. In: U. Awret & D. Chalmers (Eds.), *The Singularity: Could artificial intelligence really out-think us (and would we want*

it to)? (pp. 33-109). Andrews UK Ltd. ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=4770250

Chin, J. K., Hagstroem, M., Libarikian, A., & Rifai, K. (2017, July). *Advanced analytics: Nine insights from the C-suite*. McKinsey Analytics. Retrieved 9 September, 2020 from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Advanced%20analytics%20Nine%20insights%20from%20the%20C-suite/Advanced-analytics-Nine-insights-from-the-C-suite.pdf

Crawford, K. (2013, April 1). The hidden biases in Big Data. *Harvard Business Review*. Retrieved from https://hbr.org/2013/04/the-hidden-biases-in-big-data

Das, J. P. (2004). Theories of intelligence: Issues and applications. In: G. Goldstein & S. R. Beers (Eds.), *Comprehensive Handbook of Psychological Assessment: Intellectual and Neuropsychological Assessment, Vol. 1* (pp. 5-23). Hobboken, NJ: John Wiley & Sons.

Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Davenport, T. (2018). *The AI advantage: How to put the Artificial Intelligence revolution to work*. Cambridge: MIT Press.

Domo. (2019). Data Never Sleeps 7.0 [Infographic]. Retrieved August 17, 2020, from https://www.domo.com/learn/data-never-sleeps-7.

Dreyfus, H. (1974). Artificial Intelligence. *The Annals of the American Academy of Political and Social Science, 412*(1), 21-33.

Eisenstein, C. (2007). *The ascent of humanity: Civilization and the human sense of self*. Berkeley, California: Evolver Editions.

Eisenstein, C. (2011). *Sacred economics: Money, gift & society in the Age of Transition*. Berkeley, California: Evolver Editions.

Eisenstein, C. (2013). *The more beautiful world our hearts know is possible*. Berkeley, California: North Atlantic Books.

Eisenstein, C. (2018). *Climate: A new story*. Berkeley, California: North Atlantic Books.

Eisenstein, C. (2020, March). *The Coronation*. Retrieved 13 September, 2020 from https://charleseisenstein.org/essays/the-coronation/

Ernst, E., Merola, R., & Samaan, D. (2019). Economics of Artificial Intelligence: Implications for the future of work. *IZA Journal of Labor Policy, 9*(1), 1–35.

First, D. (2018). Will big data algorithms dismantle the foundations of liberalism? How the emergence of recommendation algorithms will shape the pursuit of happiness in the 21st century. *AI & Soc, 33*, 545-556.

Fleming, P. (2019). Robots and organization studies: Why robots might not want to steal your job. *Organization Studies 2019, 40*(1), 23–37.

Flynn, J. (2006). Towards a theory of intelligence beyond g. *Behavioral and Brain Sciences, 29*(2), 132-134.

Flynn, J. R. (2018). Reflections about intelligence over 40 years. *Intelligence 70*(2018), 73–83.

Fridman, L. (Host) (2020a, March 26). Nick Bostrom: Simulation and Superintelligence (No. 83) [Audio podcast episode]. In *Lex Fridman Podcast | Artificial Intelligence (AI)*. Lex Fridman. https://lexfridman.com/nick-bostrom/

Fridman, L. (Host) (2020b, June 22). Ben Goertzel: Artificial general intelligence (No. 103) [Audio podcast episode]. In *Lex Fridman Podcast | Artificial Intelligence (AI)*. Lex Fridman. https://lexfridman.com/ben-goertzel/

Future of Life Institute. (2015, January). *An open letter: research priorities for robust and beneficial Artificial Intelligence*. https://futureoflife.org/ai-open-letter/?cn-reloaded=1

Geraci, R. M. (2010). *Apocalyptic AI : Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality*. Oxford: Oxford University Press.

Goertzel, B. (2007). Human-level artificial general intelligence and the possibility of a technological singularity. *Artificial Intelligence, 171*(18), 1161-1173.

Goertzel, B. (2013). Artificial General Intelligence and the future of humanity. In: M. More & N. Vita-More (Eds.), *The transhumanist reader: classical and contemporary essays on the science, technology, and philosophy of the human future* (pp. 128-137). John Wiley & Sons. Retrieved 22 August, 2020 https://ebookcentral.proquest.com/lib/kbhnhh-ebooks/detail.action?docID=1138989.

Gourévitch, A., Fæste, L., Baltassis, E., & Marx, J. (2017). Data-Driven Transformation. *BCG Perspectives*, *5*.

Hanson Robotics. (n.d.). *Sophia*. Retrieved 14 September, 2020 from https://www.hansonrobotics.com/sophia/

Hao, K. (2019, January 21). *AI is sending people to jail—and getting it wrong*. MIT Technology Review. Retrieved from https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai

Harari, Y. N. (2016). *Homo deus: A brief history of tomorrow*. London: Harvill Secker.

Harari, Y. N. (2017). Dataism is our new God. *New Perspectives Quarterly, 34*, 36-43.

Harrison, N., & O'Neill, D. (2017). If your company isn't good at analytics, it's not ready for AI. *Harvard Business Review*. Retrieved from https://hbr.org/2017/06/if-your-company-isnt-good-at-analytics-its-not-ready-for-ai

Harvard Business Review. (2019). *An executive's guide to real-world AI: Lessons from the front lines of business*. Retrieved 6 August 2020 from https://hbr.org/resources/pdfs/comm/red%20hat/AnExecutivesGuidetoRealWorldAI.N.pdf

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016, December). *The Age of Analytics: competing in a data-driven world*. McKinsey Global Institute. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/The%20age%20of%20analytics%20Competing%20in%20a%20data%20driven%20world/MGI-The-Age-of-Analytics-Executive-summary.pdf

Henke, N., & Kaka, N. (2018). Analytics comes of age: McKinsey Analytics; January 2018. *McKinsey & Company*. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Analytics%20comes%20of%20age/Analytics-comes-of-age.ashx

Hoegh-Guldberg, O., Beal, D., Chaudhry, T., Elhaj, H., Abdullat, A., Etessy, P., & Smits, M. (2015). *Reviving the Ocean Economy: the case for action - 2015*. WWF International. Retrieved from https://www.wwf.de/fileadmin/fm-wwf/Publikationen-PDF/WWF-Report-Reviving-the-Ocean-Economy-Summary.pdf

Holley, P. (2015, January 29). Bill Gates on dangers of artificial intelligence: 'I don't understand why some people are not concerned'. *Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2015/01/28/bill-gates-on-dangers-of-artificial-intelligence-dont-understand-why-some-people-are-not-concerned/

Hunt, E. (2016, March 24). Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter?

Kaplan, J. (2017). Artificial intelligence: Think again. *Communications of the ACM, 60*(1), 36-38.

Joy, B. (2000, 1 April). *Why the future doesn't need us*. Wired. Retrieved 2 September, 2020 from https://www.wired.com/2000/04/joy-2/

Kulager, F. (2019, January 30). *Kunstig intelligens vil overtage verden, har jeg hørt. Så her er, hvad der skete, da jeg fik en AI til at skrive denne historie*. Zetland. Retrieved 19 August, 2020 from https://www.zetland.dk/historie/soNVPY15-me7EKAkD-7fea0.

Kulager, F. (2020, August 24). *Jeg testede en kunstig intelligens, der siges at kunne skrive som et menneske. Og jeg tror, vi er på spanden*. Zetland. Retrieved 24 August, 2020 from https://www.zetland.dk/historie/seAAwZAg-aOZj67pz-97dac

Kurzweil, R. (1999). *The age of spiritual machines: How we will live, work and think in the age of intelligent machines*. London: Orion Business Books.

Kurzweil, R. (2007). Let's not go back to nature. *New Scientist (1971), 193*(2593), 19.

Kurzweil, R. (2008). Progress accelerates exponentially. In: M. Wallace (Ed.), *The way we will be 50 years from today: 60 of the world's greatest minds share their visions of the next half century* (pp. 33-38). Nashville: Thomas Nelson.

Lesser, R., Reeves, M., Whitaker, K., & Hutchinson, R. (2018, December 8). *Winning the '20s: A leadership agenda for the next decade*. The BCG Henderson Institute. Retrieved 9 September, 2020 from https://bcghendersoninstitute.com/winning-the-20s-a-leadership-agenda-for-the-next-decade-c6f85c42dc1f

Lorentz, A. (2013). *Big data, fast data, smart data*. Wired. Retrieved 17 March 2020, from https://www.wired.com/insights/2013/04/big-data-fast-data-smart-data/

Madsen, L. H. (2019, October 14). *I 1995 byggede jeg min egen kunstige intelligens. Her er min guide til dig, der vil forstå tidens mest overvurderede teknologi*. Zetland. Retrieved 6 August 2020, from https://www.zetland.dk/historie/seWEn09v-me7EKAkD-a3e2f

Mahidar, V. & Davenport, T. H. (2018, December 6). Why companies that wait to adopt AI may never catch up. *Harvard Business Review*. Retrieved from https://hbr.org/2018/12/why-companies-that-wait-to-adopt-ai-may-never-catch-up

Manyika, J., Bughin, J., Chui, M., Silberg, J., & Gumbel, P. (2018, October). *The promise and challenge of the age of artificial intelligence*. McKinsey Global Institute. Retrieved from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/The%20promise%20and%20challenge%20of%20the%20age%20of%20artificial%20intelligence/MGI-The-promise-and-challenge-of-the-age-of-artificial-intelligence-in-brief-Oct-2018.ashx

Manyika, J., Silberg, J., & Presten, B. (2019, October 25). What do we do about the biases in AI? *Harvard Business Review*. Retrieved from https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

Marr, B., & Ward, M. (2019). *Artificial intelligence in practice: How 50 companies used artificial intelligence to solve problems*. Chichester: Wiley.

McFarland, M. (2014, October 24). Elon Musk: 'With artificial intelligence we are summoning the demon'. *Washington Post*. https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/

NBC News (2008, September 2). *Courts weigh doomsday claims*. Retrieved 10 September, 2020 from http://cosmiclog.nbcnews.com/_news/2008/09/02/4350490-courts-weigh-doomsday-claims

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist 67*(2), 130-159.

Neuralink. (2019, July 16). *Neuralink launch event* [Video]. YouTube. https://youtu.be/r-vbh3t7WVI

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reshaping business with artificial intelligence. *MIT Sloan Management Review*. Retrieved from http://marketing.mitsmr.com/offers/AI2017/59181-MITSMR-BCG-Report-2017.pdf

Russell, B. (1996). *History of Western Philosophy*. London: Routledge. (pp. 111-120)

Sabeti, A., & GPT-3. (2020, July 31). Are humans intelligent? A salty AI op-ed. *Arram Sabeti*. https://arr.am/2020/07/31/human-intelligence-an-ai-op-ed/

Sandberg, A., & Bostrom, N. (2008). Global Catastrophic Risks Survey. *Technical Report #2008-1*. Future of Humanity Institute: Oxford University, pp. 1-5.

Satell, G. & Sutton, J. (2019, October 28). We need AI that is explainable, auditable, and transparent. *Harvard Business Review*. Retrieved from https://hbr.org/2019/10/we-need-ai-that-is-explainable-auditable-and-transparent

Silberg, J., & Manyika, J. (2019). *Notes from the AI frontier: Tackling bias in AI (and in humans)*. McKinsey Global Institute. Retrieved from

Smith, T., Stiller, B., Guszcza, J., & Davenport, T. (2019). *Analytics and AI-driven enterprises thrive in the Age of With: The culture catalyst*. Deloitte Insights.

Sophia the Robot. (2020, January 14). *Sophia the Robot gives a glimpse of what's to come in 2020 #Sophia2020* [Video]. YouTube. https://youtu.be/WPhPrnw4QKk?t=90

Stenvik, B. (2020, July 4). *Zuckerberg og Musk vil lede menneskeheden ind i en science fiction-fremtid. Her er, hvorfor vi roligt kan sige nej tak.* Zetland. Retrieved 19 August, 2020 from https://www.zetland.dk/historie/s81ELg4Z-me7EKAkD-d9c6a

Sternberg, R. J. (2013). Intelligence. *WIREs Cognitive Science, (3)*, 501-511

Søndergaard, K. (2020, August 21). Årets EISA-vindere er kåret: Her er verdens bedste tech-produkter. *Jyllands-Posten*. Retrieved from https://jyllands-posten.dk/livsstil/digitalt/ECE12352531/aarets-eisavindere-er-kaaret-her-er-verdens-bedste-techprodukter/

Tegmark, M. (2018). *Life 3.0: Being human in the age of artificial intelligence*. London: Penguin Books.

Vincent, J. (2020a, July 30). *OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws*. The Verge. Retrieved from https://www.theverge.com/21346343/gpt-3-explainer-openai-examples-errors-agi-potential

Vincent, J. (2020b, August 18). *Facebook and NYU use artificial intelligence to make MRI scans four times faster*. The Verge. Retrieved from https://www.theverge.com/2020/8/18/21373335/faster-mri-scans-ai-machine-learning-facebook-nyu-research-clinical-study

Wang, F., Zhang, J. J., Zheng, X., Wang, X., Yuan, Y., Dai, X., Zhang, J., & Yang, L. (2016). Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, *3*(2), pp. 113-120.

Yudkowsky, E. (2008). Artificial Intelligence as a positive and negative factor in global risk. In: N. Bostrom & M. Ćirković (Eds.), *Global catastrophic risks* (pp. 308-345). Oxford: Oxford University Press.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for the future at the new frontier of power*. London: Profile Books.