1

Introduction

It is possible that in the distant future we will look back at human history and speak of three major disruptive technological innovations. The transition from a hunter-gatherer culture to a sedentary agrarian culture with animal husbandry in the Neolithic Age, the transition to the machine age based on fossil fuels in the nineteenth century, and finally the digital revolution of the twenty-first century: the use of artificial Intelligence[1]. If this is the case, we are only at the beginning of a technological revolution, similar to what Europe experienced in the first decades of the nineteenth century. And just as then, technological innovations today are accompanied by apocalyptic fears, but also by euphoric expectations.

This book deals with the cultural and philosophical aspects of Artificial Intelligence and pleads for a digital humanism. Digital Humanism is technology-friendly, but also human-friendly. It sets itself apart from the apocalyptics because it trusts human reason, but it also sets itself apart from the euphorics because it respects the systematic limits of digital technology.

The dream of the creation of artificial beings has been part of mythological narratives for thousands of years. In antiquity, it is the myth of Prometheus, a god from the Titan family, who creates thinking and feeling clay beings without divine permission and is bitterly punished by Zeus for it. In the Middle Ages, we find the story of the Golem, an artificial being made of clay, which is mute and not capable of reason, but possesses great strength and can carry out orders. Literature also uses the myth of the artificially created being. In the story "The Sandman" by E. T. A. Hofmann (1816) the protagonist falls deeply

[1] If we write "artificial intelligence" we implicitly accept the existence of artificial intelligence. If we leave it open we should capitalize "Artificial Intelligence" which we do in this book.

J. Nida-Rümelin, N. Weidenfeld, *Digital Humanism*,
https://doi.org/10.1007/978-3-031-12482-2_1

in love with an animated doll named Olimpia and in the course of events ultimately loses his mind over it. Perhaps the most famous example from this period is Mary Shelley's novel *Frankenstein* or *The Modern Prometheus* (1818). In this tragic story, a Swiss scientist creates an artificial human. This artificial man arouses so much disgust and fear due to his size and ugliness that he cannot connect with human society and, on the contrary, accumulates more and more rage and hatred within himself. In the end, he kills the bride of his creator and himself.

Today we might call our contemporary humanoid robots "frankensteins" had there not been the play *R.U.R.* by Czech writer Karel Čapek in 1920. This drama is about a company called R.U.R. (Rossum's Universal Robots), which produces artificial humans called "robots" and abuses them as cheap laborers, who, however, in the course of the story rebel against their slavery and wipe out humanity.

In the twentieth and twenty-first century, the robots live mostly in sci-fi novels such as the ones by Stanisław Lem[2] or of the US-American author Philip K. Dick.[3] In recent years, US-American sci-fi blockbuster films have heavily drawn on the mythological figure of the artificial human, which now appears as a robot that cooperates with humans on earth and on spaceships.

Apart from these, there is also the idea of a fully digitalized world which sci-fi films and novels have taken up. The vision is almost always dystopian: there are worlds completely dominated by machines like in the film *The Matrix* (directed by Wachowksis, USA, 1999) or futuristic nightmarish societies such as the one in the film *Demolition Man* (directed by Marco Brambilla, USA, 1993), in which people act and interact based on digital instructions and even sexual contact may only take place through the mediation of digital media.

In the meantime, many things which were fantasized about in the history of mankind have become reality, the most famous example being Captain Kirk's foldable "communicator" from *Star Trek*, which was technologically realized some 50 years later in the form of the StarTAC mobile phone by Motorola.

It even seems that the myths merely take on a form impregnated by new technologies but remain unchanged at their core. The myth of the machine in human form that takes over in the end, the myth of the animated doll, the myth of a friendship between man and machine. But unlike previous centuries, these myths now appear to be revitalized by concrete technological options.

---

[2] See, for example, *The Star Diaries* (1976) or *Golem XIV* (1985).

[3] The US author Philip K. Dick wrote numerous books and short stories on which many US films such as *Blade Runner, Minority Report*, or *Total Recall* are based.

There can be no doubt, we are living in a time of technological upheaval. This century and the next—many are convinced—will be the age when robots will have taken over many kinds of human work. They will deliver parcels, produce goods in factories, drive taxis, act as bank advisors, explore space, work in call centers, operate alongside doctors in hospitals, and possibly write novels and create art. But one need not to look so far ahead. Digitalization has already permeated not only our working world but also our private lives and has had a great influence on our cultural, political, economic, and social life.

This development raises many questions as to what the consequences will be for us. Some, such as bestselling authors Daniel H. Wilson (*Robopocalypse*, 2011) (a former research associate at Carnegie Mellon University who earned a PhD in robotics) or Stephen Hawking[4] or philosophers such as Nick Bostrom[5] warn us that robots will one day surpass the human species in thinking and acting abilities and turn against humanity.[6] Others, harbor utopian hopes for a new, digital world in which digital robots as modern slaves perform human work and establish a realm of unprecedented freedom for us.

There is much to suggest that what is called "strong AI" in the digitalization discourse, i.e., the thesis that software systems have consciousness, make decisions, pursue goals, that their performances are not merely simulations of human abilities but realize them, will one day be considered a form of modern animism, i.e., the ensoulment of the non-ensouled, which can be seen as a regression into childlike modes of interpretation.

Of course, such a digitalization ideology does not present itself as regressive and childish, but on the contrary as rational and scientific. It has a long cultural history. It begins in our cultural sphere with the Pythagoreans in the fifth century BC. It is the idea of a world strictly ordered in numerical relations, the harmony and rationality of which is only revealed in mathematical analysis. Two hundred years later, the Stoic philosophers added to this theory the thesis of the correspondence between world reason and human reason (logos). According to this theory, people are only able to think and act rationally because they can participate in world reason. The logos orders the world

[4] Stephen Hawking warns in many interviews against unrestrained use of Artificial Intelligence. For example, in the Focus interview in 2015: "Our future is a race between the growing power of technology and the wisdom with which we use it." Online at: http://www.focus.de/wissen/technik/wird-man-sie-kontrollieren-koennen-stephen-hawking-warnt-in-100-jahren-sind-computer-intelligenter-als-menschen_id_4681638.html (last accessed on 16 April 2017).

[5] Bostrom (2014).

[6] German scientists such as the German philosopher Thomas Metzinger also warn of the negative effects of an "AI arms race," at the end of which super-intelligent software could emerge that detaches itself from its computer and, like a large, uncontrollable virus, activates itself in ever new places and globalizes itself and its targets.

according to strict deterministic laws and human beings have to fit into this world reason. Even the Stoics and their opponents noticed, however, that there is a tension between a world view of comprehensive determinism and a view of oneself as a free and responsible human agent. If AI ideology leads to a new edition of this conflict, then digital humanism overcomes this conflict.

In this book, we develop the main features of a digital humanism as an alternative to what can somewhat simplistically be called "Silicon Valley ideology." Silicon Valley ideology related to the original American, Puritan hope of salvation, of creating a world of the pure and righteous who have left filth and sin behind. Its central values are transparency and predictability, economic success and patronage. In times of digital transformation this included the dream of a perfectly constructed digital counterparts whose construction excludes any error leading us into a technological utopia. The key concept here is that of Artificial Intelligence, charged with implicit metaphysics and theology, a self-improving, hyper-rational, increasingly ensouled system whose creator, however, is not God but software engineers who see themselves not merely as part of an industry but of an overarching movement realizing a digital paradise on earth based on transparency, all-connectedness, and non-ambiguity.

The Silicon Valley ideology takes humanist impulses as its starting point, only to transform them into anti-humanist utopias. It begins with the improvement of the human and ends in its final—and inhumane—overcoming. By wanting to improve human life on the planet, it starts to question the conditions of humanity. In the course of this, humanism is thus being transformed into transhumanism, leading to a technicist utopia in which the human is left behind. Digital humanism opposes this and offers instead a new ethics for the age of Artificial Intelligence.