Údaje na deskách fyzické kopie obsahově krom jména vedoucího a názvu oboru identické

CHARLES UNIVERSITY

FACULTY OF HUMANITIES

SHV Programme – Philosophical Module

Matyáš Mandík

AI´s Impact on the Environment and Anthropocene

Bachelor's thesis

Supervisor: Juraj Hvorecký, Ph.D.

Prague 2024

FORMAL DECLARATION

I hereby declare that I have written this diploma thesis solely by myself and I agree with its eventual publication in print or electronic form. All sources and literature were properly cited. The work was not used to obtain different or the same title.

……………………………………………..

In Prague, the Author´s signature

ACKNOWLEDGEMENT

I wholeheartedly thank the supervisor of this work Juraj Hvorecký, PhD. Namely for his willingness and highly valued guidance that I have received from him. It is thanks to him that I was able to not only compose a thesis that corresponds with my interests in the academic field, but I was also enriched with new experiences and knowledge through his patient guidance. Last but not least, many thanks also go to my high school teacher and friend Mgr. Josef Makoč, who introduced me to philosophy and the beautiful potential it has for the human mind and understanding our relation to the world around us.

Abstrakt:

Tato práce filosoficky prozkoumává možnosti vlivu a dopadu, který má umělá inteligence na globální životní prostředí v současném období antropocénu. Tento dopad a vliv pak rozebírá ve dvou dichotomických rovinách. Jedné pozitivní, která v sobě skýtá možnost a vizi lepší budoucnosti v kontrastu s jejím protikladem, který se odehrává nyní; ten, pokud bude opomenut a neartikulován, vnese do budoucí historie lidí a Země ještě menší naději na zvrácení již tak naléhavé situace životního prostředí. Toto enviromentální zaměření je protkáno etickými otázkami týkajícími se lidí a komunit, které umělá inteligence ovlivňuje, a zejména kterými je ovlivňována a využívána pro osobní prospěch.

Abstract:

This thesis, philosophical in nature, explores the influential and impactful possibilities that artificial intelligence poses upon the environment in today's epoch of the Anthropocene. Said influence and impact are being deconstructed in two dichotomous ambits. One, positive and comprising a possibility and vision of a better future in contrast with its opposition, which is happening as we speak – it, if forgotten and inarticulate, will sow into the future history of people and Earth a still smaller hope of reversing the already dire state of the environment. This environmental focus is interwoven with ethical questions concerning people and communities that artificial intelligence influences, and those who influence artificial intelligence and use it for their desires as well.

Klíčová slova:

Umělá inteligence, antropocén, životní prostředí, etika, pravidla, aplikované užití, vývoj, udržitelnost, zodpovědnost, moc.

Keywords:

Artificial Intelligence, Anthropocene, environment, ethics, policymaking, applications, development, sustainability, responsibility, power.

Table of contents

This page is intentionally left blank.

Introduction

This thesis discusses the vast and not-so-new topic of artificial intelligence in a philosophically, yet not solely philosophically, oriented way of thinking. It namely deals with the impact and its many forms that artificial intelligence has on the planetary environment, people, and the conjoint future. Though things are rarely that simple and straight, the assumed influence of artificial intelligence does not only go toward us and the environment, as we can also eventually reshape artificial intelligence and or the environment, and vice versa.

As said, my work aims to show how artificial intelligence influences the environment, and in turn people, through its extensive usage in various sectors including, for example, surveillance, facial and affect recognition devices, object recognition, and labelling. All of the above is possible only thanks to a murky, rarely talked-about world of the physical structure that all artificial intelligence systems have in common. Their physical structure comes in many different shapes and forms varying from extensive human labor, supply chains, and material extraction. It is mainly thanks to these hidden physical qualities that artificial intelligence influences our environment so negatively.

To look at the bright side there are many possible outcomes in which artificial intelligence can help humanity towards reverting, nullifying, or simply just slowing down the ongoing global climate crisis. Though some of them are purely speculative, there are also some examples and ideas that are already based on a solid ground basis.

The methods and data used in my thesis are both qualitative and quantitative. As this work has a strictly theoretical approach, I use solely secondary data drawn from researchers cited in my work. The reasoning behind the use of this methodology is due to it being a standard in philosophical works; as well as taking into account the fact that I aim at presenting and highlighting an already existing problem that, in my eyes, needs further addressing. Additionally, I aim to counter the matter of possibly biased or inaccurate sources by drawing on multiple sources for subsequent comparison; if the source data aren´t correlative with each other, I then tend to state an estimate drawn from the already mentioned samples.

My motivation for writing this thesis draws from my philosophical interests combined with creating an applicable solution, or at least proposition, that has a lasting impact and real use in our world and my innovative nature. As I strive to earn a doctorate in my academic field eventually, I deem this subject to be an eligible foundation that can be built upon for my future interests in research. I also hope my work can contribute and positively impact the world and the people living in it.

1 What exactly is “AI”?

Today, thanks to the late boom of its heavy commercial use, we all know what the abbreviation “AI” stands for, or at least imagine some of its foundational functions. However, it is a rather ambiguous term as its deeper meaning and functions vary, depending on the circumstances or situations that are referred to in various fields of application. Because of this fact and as a way of preparing a firm ground regarding its meaning for this thesis, allow me to give a short, but needed introduction to the history of artificial intelligence and its meaning; as well as their explanation and contextual embedding to my framework, behind the many terms related to it. I deem it overall helpful to have a firm grasp of a subject of one’s interest, especially given the vast context in which artificial intelligence operates; it also gives You, the reader, a better understanding of how I, the author of this work, think of the field of artificial intelligence, thus leading to a better mutual understanding. I am to consciously use a certain generalization: because just as an IT technician will see artificial intelligence more in terms of code and software, a philosopher will deem it ethically interesting; a sociologist, on the other hand, will perhaps see a great tool for methodology or measurement.

The very essence of creating artificial or better beings, resembling today’s artificial intelligence, can be traced back to antiquity, both in the Greek and Roman worlds. As Fron (2019) observes, ancient perceptions of automata[1] were based on a different fundamental perception from the one we have today – looking for utopian features meant to look into the ever-evolving and glorified past based on mythos combined with actual history rather than the future. These visions, combined with mythical descriptions, of anthropomorphic robots all reach the same conclusion that human creativity and invention are limited to the realm of mere idealized imitation of life itself, and the actual creation of a genuine spring of creativity can only be conducted by divine powers (Ibidem, 1-3).

Medieval perception of automata was heavily influenced by the continuous tradition of antiquity; however, it is also important to understand that it represents a significant shift from it. The earliest works of engineers, creating moving dolls or robot-like figurines, can be dated to the 9th century AD. Yet, any mechanical knowledge in most of Europe was lost until the

[1] A “machine which resembles and is able to simulate the actions of a human being” (OED, 2023).

fourteenth century, as technical knowledge was associated with magic and wizardry. Important books often show sketches with descriptions of complex mechanical figures or humanoids for various purposes often associated with the work *servants* did. There are even a few preserved and functional automated figures from the 16th century with simple spring and cog mechanisms (Ibidem, 4-7).

Modern perceptions of automated machines differ even more, mostly because of the influence of a late 18th-century machine called the mechanical or chess Turk. The machine was a brilliant fake, with a chess grandmaster hiding inside of it - and yet it amazed everyone on its tour through Europe and United States. Apart from popular amazement, there was also fear, as this robot seemed to have human-like intelligence, and thus challenged the core of humanity, sparking fear behind every future doubt and criticism of modern robotics. Nowadays, computers are the modern successors of the automaton, yet they still have an element of trickery to them, mainly in the form of intense specialization, pre-programmed answer alternatives, or intense human intervention (Ibidem, 7-9). It is important to keep in mind that the inspirations from antiquity only make sense to the point where people think that intelligence was tied to our physicality – in the end, it only started to make full sense when the human brain became the substrate of our intelligence in the early part of the 20th century. We therefore have the answer as to why we moved from artificial automated slaves towards chess-playing robots: we wanted to capture the intelligence required for such complex tasks, and hence the shift from the original antique inspiration.

Muthukrishnan (2020) gives a vague and brief answer to our question about the broader meaning of artificial intelligence. He says that "artificial intelligence (AI) is revolutionizing many industries by performing tasks that typically require human intelligence to solve" (Ibidem, 393). Revolutionary changes take the form of enhancement, simulations, and augmenting the intelligence of humans with usage in "[…] banking, conversational bots used in customer service and precision diagnostics in health care" (Ibidem) and much more, as we shall see. To truly catch a glimpse of the technical complexity of artificial intelligence, we need to take a look at its fields and subfields. The most notable examples that Flasiński (2016) talks about in his work *Introduction to Artificial Intelligence* are natural language processing, computer vision, reinforcement learning, speech and pattern recognition, machine learning, deep learning, and many more specialized ones (Ibidem, 16-27). Just as human intelligence has multiple layers, based on cognitive functions and other perceptions that make up our understanding and perception of reality and the world around us, artificial intelligence can also be decomposed similarly. One important difference is that while human intelligence, or at least its majority, always has all (or most of) its tools available and cooperates as one, these subfields of artificial intelligence can operate independently or they can be combined in various ways, depending on the task we want it to perform. In a simplified example, if you wanted an algorithm that recognizes speech patterns as well as identifies facial expressions, you would skillfully combine the subfields of language processing, speech recognition, deep learning, and computer vision.

According to Muthukrishnan (2020), machine learning creates algorithms that can learn complex tasks and develop predictive models based on vast sample data. Along with deep learning, a subfield of machine learning that processes data with maximum optimality, machine learning, and deep learning were both initially designed to imitate the human brain's neural activities. Starting as mere simulations that aimed to imitate the function of a singular neuron, they grew into the founding layers of deep learning we know today (Ibidem, 393-94). Crawford (2021) brings forth an important point by noting that "in the field of artificial intelligence, where the belief that human intelligence can be formalized and reproduced by machines has been axiomatic since the mid-twentieth century […] AI systems have repeatedly been described as simple but humanlike forms of intelligence" (Ibidem, 5). Flasiński (2016) calls these attempts at recreating human intelligence "evolutionary computing" – a paradigm in early days of AI field which stresses the principle that "simulating nature at its biological layer is a proper methodological principle for constructing AI systems" (Ibidem, 11) - In other words, it aimed at recreating both the brain's anatomical, physiological, and corresponding psychological elements. As Muthukrishnan (2020) puts it "The 1956 Dartmouth conference is generally considered the moment that AI formally recognized and obtained its name and its mission. This conference […] may be considered the 'birth of' the field of AI, including the assertion that 'every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it'" (Ibidem, 394-95). Crawford (2021), for example, defines the mentioned evolutionary computing as "one of the core disputes in the history of artificial intelligence" (Ibidem, 5) that was heatedly discussed at a following Dartmouth conference in 1961 and had polarizing conclusions. Most of the greatest computer scientists at the time argued that the difference between human intelligence and that of machines is illusionary. This led to a counterargument from the philosophy professor Hubert Dreyfus stating that “the assembled engineers ‘do not even consider the possibility that the brain might process information in an entirely different way than a computer’“ (Ibidem, 6). Since then, there have been a lot of changes both in concentration on newer machine learning techniques, which I will discuss later, and the disputes over the capabilities of AI that at the time halted due to the computational power of newer hardware components, as we will shortly identify in the following historical window of how the technical approaches of artificial intelligence established themselves.

Although the field of artificial intelligence formulated its newly set priorities through the Dartmouth conference and seemed, research-wise, on a good path regarding some real progress, Muthukrishnan (2020) goes on to describe two AI winters – “an overall decline of interest in the field by investors” (Ibidem, 396). The first AI winter took place from 1974 to 1980 and was caused by unrealistic expectations from the public, investors, and in some cases even the researchers themselves. Unfulfilled expectations led many governmental bodies to stop funding the research. This caused most researchers to move from algorithms learning representations of data towards expert rule-based systems (Ibidem). As Flasiński (2016) observes: “Although such negative opinions of eminent AI authorities caused limited financing of research into neural networks, it continued in the 1970s and 1980s and resulted in many successful results” (Ibidem, 10). Going back to Muthukrishnan (2020) the first AI winter ended by the end of the 1980s thanks to concepts and publications that once again sparked the interest of investors. And yet, the second AI winter came only a couple of years after this surge and lasted up until the mid-1990s due to the “increased hype in the capabilities of neural networks without sufficient advancement in computing power” (Ibidem, 396). To put it more simply, these theories were too advanced for their time and the machines, along with their hardware, that were supposed to support the implementation of these theories could not do so, because they lacked the power output, and the AI field once again came to a halt.

“The interest in AI resurged toward the mid-1990s as computational power increased and could support the development of neural networks. The microcomputer revolution and

Moore's law[2] both describe the advancements computers had in this decade alone that allowed the replacement of the traditional machines that were throttling AI development. The capabilities of AI paired with sufficient computational power as demonstrated in 1997 when IBM developed the chess-playing supercomputer Deep Blue. Deep Blue defeated the chess champion Kasparov, which led to many publications and documentary films that attracted the public's attention to the field once again" (Ibidem, 396). Then, roughly during the last decade, two important factors, allowing further progress spike in the field of artificial intelligence appeared, namely "data storage" and "graphical processing units (GPU)". Data storing became easier, thanks to the reduction of physical space, cloud storage technologies, and lower costs; the performance and overall capabilities of GPUs improved as well, allowing for better power output as the limitations of previous computational hardware were overcome. One of the breaking points in the field of AI of this era was the shift back to deep learning caused by the invention of AlexNet, a revolutionary program that won the ImageNet Large Scale Visual Recognition Challenge – a dataset of 1.2 million images and 1000 categories in which the participants aimed at correctly recognizing the images and categories via their programs. This led to the creation of many networks that were based on deep learning and numerous experts continue to build on them to this day (Ibidem, 396-98). One such example is ImageNet or ChatGPT (although they are both fundamentally very different) which we will discuss later in this thesis. It is also crucial to note, for clarity's sake, that ChatGPT functions on an unsupervised network learning principle, whereas the examples of AlexNet or ImageNet run on a supervised network. These terms are pretty much self-explanatory, as one has a predefined dataset it learns from and the other does not (same applies to information etc.), yet the most important thing to know is that the internal structures differ, and the computational field and science have moved from unsupervised to supervised network learning systems and back again.

To summarize this brief history overview of how artificial intelligence came to be, we can say that from the 1950's onwards, progress in the field was steady and this first period was in the spirit of semantic artificial intelligence, where everything is done through explicit

[2] An observation made in 1956 by Gordon Moore, later named Moore´s law, that the number of transistors in an integrated circuit doubles about every two years. It is an observation and projection of historical trends that surprisingly corresponds to developments until now (Schaller, 1997, 52).

programming commands. The "AI winters" correspond to the depletion of this semantic approach and the move towards neural networks. As hardware computing power rose, the semantic approach returned to full strength and fulfilled its theoretical potential, as the new hardware that supported it enabled previously impossible results. The most important turning point came in the 1990s when then top-notch hardware was being combined with neural network approaches, which resulted in the unprecedented possibility of simulating the human brain on a previously unimaginable scale. To underline this a bit more, even though the algorithms haven't changed much since the 1950s, we can now, thanks to the computational power of today, work with them on a much larger scale.

Today, AI is "pervasive, often invisibly embedded in our day-to-day tools and as part of complex technological systems" as Coeckelbergh (2020, 3) puts it. This invasion of our domains, which we often consider private, is possible mainly because of the enormous data volumes that can be found on the internet and which we all help create through our usage of ever-present smartphones and other such gadgets[3]. This, in turn, allowed the mass use of algorithms that have taken over activities such as planning and decision-making trained on our data. As of this moment, artificial intelligence has great applications in "transport, marketing, health care, finance, and insurance, security and the military, science, education, office work, and personal assistance, the arts, agriculture and of course manufacturing" (Ibidem, 3-4).

This brings me to my understanding of artificial intelligence and what background I imagine when using this complex term, as we have briefly touched on in the first lines of this chapter. First and foremost, as we have seen, the idea of automata, a humanoid robot, or simply just something close to human-like intelligence has been with us for as long as our ancient roots go; slowly evolving and taking shape into the AI we now have. What I mean to emphasize is that we are not dealing with the latest technology or a viral phenomenon – as we shall see later in this thesis, the practices of making and creating artificial intelligence go way deeper into our history than we realize and are rather problematic when put face to face with society, environment, and ethics.

Secondly, despite its misleading image of mental domain, artificial intelligence has a structured physical basis, as we shall observe in greater detail through a discussion about this

[3] The cause of this "invasion", as well as its roots will be discussed later on in greater detail.

physicality, that enables it to function, and which has to be built using real-world materials. Thirdly, as much as artificial intelligence can be helpful and potentially even greatly enhance humanity as a whole, it can also bring irreversible damage and be seized by a handful of individuals for their benefit – an important duality to keep in mind. Lastly, I would like to take a stand against futurism, speculations about the possible future, and escapism[4] that many public figures, including Elon Musk as a prime example, venture into. My work is not to be taken as a speculative work of what possible futures look like, but rather current reality-based research of our most pressing matters that we are to deal with in the near future and how to overcome them with the tools currently available to us. In other words, I aim at it to be a useful inspiration, not a dystopian or utopian tell-tale of the times to come, as it is quite common when discussing the subject of artificial intelligence and environmental crisis alike. The future of artificial intelligence should shift from the current futuristic speculations of it either enslaving for us, or us, towards how we can align these systems with the current goal humanity should have – saving our planet and very civilization.

[4] In this case not only the traditional meaning of turning your attention away from more pressing matter in order to distract one self, but a more literal take that consists of abandoning our planet as it is doomed anyways and seeking a new home planet to terraform.

2 Current environmental global situation

Before discussing our central topic further, we must lay out and thoroughly explore the current environmental situation we find ourselves in. Kolbert (2014) tells the story of a likely Sixth Extinction, which begins with the emergence of a new humanoid species around two hundred thousand years ago. Its members are not the most unique or particularly strongest, yet they steadily grow in population, move into new territories with various life forms and landscapes, and quickly adapt and innovate. Homo sapiens, as we later begin to call ourselves, then spread to almost every habitable corner of the world and begin to reproduce at a very rapid phase. As our population exponentially grows, we cut down forests and displace other organisms as we see fit to feed and sustain ourselves. As time goes on, humans discover new forms of energy, hidden in Earth's crust and they begin to change the very composition of the atmosphere, which creates a chain reaction that results in alteration of climate and oceans. Although similar events, known as the Big Five, occurred on our planet previously and wiped out significant ratio of life on Earth, no creature ever altered life itself as much as humans did, and continue to do so (Ibidem, 1-3). As of today, we may very well be past the point of return concerning the mass extinction of plant and animal life alike caused by human activity and correlated to the worsening state of the environment.

As Nadakavukaren (2020) aptly describes this situation: "[…] people often deceive themselves into believing they are all-powerful creatures, apart from the rest of nature. Yet […] humans, like all other living organisms are inextricably bound up in the web of interdependency and interrelationships that characterize life on this planet. Human health, well-being, and indeed survival are ultimately dependent on the health and integrity of the whole environment in which we live" (Ibidem, 2). Despite this dependency on the world around us, we continue to damage and alter it; what's worse, we do this with full consciousness of the negative impact our actions have. Wallace-Wells (2019) observes an estimated 31 % increase of carbon in the atmosphere than there ever was in the span of the last

800,000 to possibly 15 million years[5]; more than half of the carbon released into our atmosphere through burning fossil fuels originated in the last thirty years. We have therefore done more damage to the future of our planet in the last three decades than we did in the past many millennia before it – and that's only accounting for our carbon footprint, which is only one of many ways we continue to damage the environment we are dependent upon (Ibidem, 2).

In 1958, when Hanah Arendt wrote her first edition of *Vita Activa* (2009), she very interestingly contemplated the relation of humans to their environment as unmoored from its earthly conditions, because humans are no longer creatures bound to Earth in the same sense as every other creature is – humans, in the sense of living beings, are trapped in the realm of physical living which's nature is in the form of dependence, and yet they actively recede towards a meretricious world created by themselves, moving their foundations further and further from the Earth, towards Cosmos (Ibidem 9-19). Four years after *Vita Activa* was published, Nadakavukaren (2020) identified a book called *Silent Spring* (written by Rachel Carson) as a turning point in the public view on environmental damage (Ibidem, 3). As she says: "[…] the book's publication marked the opening salvo of the environmental revolution (in fact, Carson was the first to popularize the word 'environment')" (Ibidem). The book's success merited in the fact that it made the wider public see the real extent of a problem they had largely ignored up until then. Before the 1960s, the preservation of wilderness or conservating resources came only in the early 20th century with origins in America, where the destruction of unspoiled nature accelerated after the Civil War in the limited forms of statutory amendments and small local regulations thanks to Theodore Roosevelt, John Muir, Gifford Pinchot, and other early conservationists. After *Silent Spring*, even the public acknowledged and became aware of the issues of toxicity, polluted rivers, and smoggy skies, which resulted in the demand for policy actions from the hands of the US government. The environmental movement then grew, paradoxically supported by a series of eco-disasters on a global scale in

[5] The Intergovernmental Panel on Climate Change (IPCC) presents numbers in roughly the same range with the major deviation being the estimated between 800,000 to 15 million years, whereas IPCC states 420,000 to 20 million years (IPCC, Solomon et. al., 2007); but with the time window of 12 years, we can expect a difference given the technological and methodological advance made in the meantime, which allows for more precise measurements.

the 1970s which helped spread the message to many communities around the world (Ibidem, 2-4).

Of course, environmentalism did not originate solely in America, nor did it spread only because of a singular book. It has deep roots, local and global, that each vary in their approach towards how they view preserving Nature or environment and at what scale. They also differ in size - local neighborhood associations on one end, and global multibillion-dollar associations with corporate structures led by experts in their respective fields (Gottlieb, 1993, 3-4). We can also find plenty of different outlooks for the future, some being very optimistic, and some chiming a death bell for all life on Earth. Regardless of the time we have before that happens, a great majority of works dealing with today´s global environmental situation agree that the effects of climate change will have in one way or another catastrophic consequences for the planet and its life forms; it is to be noted that world is not only black and white, and so there is also a widespread shared opinion that solutions to this dire situation do exist and are, to some degree and forms, attainable.

David Wallace-Wells through his book Uninhabitable Earth (2019) represents one of many alarming voices, arguing that climate change is happening now, faster than we think, and not only at the level of sea rise and the melting of remote Arctic caps – it is a crisis of the natural world, as well as the human one; no corner of the Earth will be left untouched and no lifeform unchanged (Ibidem, 1-3). He also adds that "[…] If the planet was brought to the brink of climate catastrophe within the lifetime of a single generation, the responsibility to avoid it belongs with a single generation, too. We also know that second lifetime. It is ours" (Ibidem, 4). Nadakavukaren (2020) goes on to describe the difference in perception of environmental problems in the 1970s and nowadays through the meaning of a peculiar kind of disassociation and ignorance; she does so thanks to the description from Gus Speth and his book *Red Sky at Morning* from 2004. Before that, pollution or problems with toxins were much easier to understand, because they were closer to each individual's home or region and presented a sort of imminent and highly visible danger. Today's problems have become a scientifically complex global agenda. On top of that, change often evokes a certain distrust – they seem too far in the future or too far geographically from our own homes to be worried about (Ibidem, 5). She then elaborates: "Similarly, since the major impacts of climate change, tropical deforestation, and species loss will be on people and places far away, rather than on our communities, we're less motivated to take potentially painful policy action than was the case four decades ago. All of these factors have diluted the once emphatic demand that the government 'do something right now!' to address these issues" (Speth, 2004 in Nadakavukaren, 2020, 5).

The impact of human influence on Earth and its life is unprecedented, ranging from deforestation, air pollution in the form of greenhouse gases, forceful removal of species of plants and animals to different biotopes, loss of biodiversity, soil salinity and overall quality, river modification, water, and thermal pollution, excavation, sediment, and waste dumping and so much more, as Goudie (2006) observes in his work *The Human Impact on the Natural Environment*. Collected from various scientific resources, his interpretation of our future is quite grim; though he acknowledges the fact that any prediction of the future is considerably uncertain as it is difficult to account for all the complex systems and reactions of the many various entities that play a role in the climate change (Ibidem, 234-235). Many of these future predictions include changes in the biosphere such as altitude changes for vegetation zones, which in turn increases desertification and decreases in forests; with the temperature continually increasing, the face of land and land forming process itself changes, driving forth melting of ice caps which results in floods and water rises; significantly more runoff water now appears thanks to factors like urbanization and mentioned dryland expansion which in turn alters the rainfall and hydrology again resulting in cumulating heating of the atmosphere (Ibidem, 232-294). Then, there are of course many more changes at the local regional level, not even mentioning the catastrophic impact all this has on species of fauna and flora, including humans. The baseline here is the utter fact that all of these phenomena are not happening in isolation and are only remotely related to each other, but that they cascade off of each other, and are closely intertwined. As far as we can tell, they form a peculiar, sort of cyclic event that has already started happening, and will only inevitably accelerate depending on our current contribution to the destruction of our environment.

Concerning the very future of humankind itself, Kolbert (2014) proposes multiple possibilities. One is the fact that we too will sooner or later succumb to the damage we have done to our biological and geochemical systems on which we are dependent (Ibidem, 253). The second is a bit more optimistic, yet naive in a way – "[…] human ingenuity will outrun any disaster human ingenuity sets in motion" (Ibidem). Whichever outcome, Kolbert identifies as the most inescapable our current unintentional decision that will forever determine which evolutionary paths will remain open or closed forever – this ability, unique to our species, will likely become our most lasting contribution to the planet (Ibidem, 254).

With this background, we see why the environment is important for a thesis focusing on the ethics of artificial intelligence and explaining a mutual relation between the two.

3 The physical behind AI

Kate Crawford, through her work *Atlas of AI* (2021) presents a unique way of seeing artificial intelligence as it "[…] is neither *artificial* nor *intelligent*. Rather, artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications" (Ibidem, 8). In the following paragraphs, I will elaborate on these physical features and material foundations of artificial intelligence in greater detail, as we need to familiarize ourselves with how they contribute to damaging our environment, as well as many human communities while being used by a few individuals owning giant tech-companies for their purposes and power.

Terms such as *cloud*, *data*, and *algorithms* are often misleading, purely functional, and floating in the air – they form the backbone of the industry of artificial intelligence; yet they cannot function without being physically built and embodied (Ibidem, 30). Our perception of the physical elements and constitution of artificial intelligence is, in my opinion, not dissimilar to an interaction between us humans; yet it remains greatly distorted. When we interact with somebody, mainly verbally, we find ourselves in sort of a metaphysical space where our minds interact and exchange information. In addition to this plane, we also acknowledge the other person as "someone". This "someone" is the originator of the information flow towards me and has physical bodily features and characteristics that we can identify. In our everyday lives and interactions, we syncretize these perceptions into one, which forms the reality of "my friend telling me how he looks forward to the opera next weekend". Even though we consciously do not do so, we know our friend consists of cells, bones, flesh, nutrients from food, and other such things that can support his physical being in this world, which in turn allows his metaphysical being, be it mind or soul, to exist: we subconsciously know that through interaction with his mind, we are interacting with his physical body and vice versa.

In the interaction between a human and an artificial intelligence system, we also enter a peculiar metaphysical space, but a rather distorted one. It takes the form of a master-slave relationship, where humans ask or demand a particular function from an artificial intelligence system of our choosing and expect a predefined result or action. At that moment, our perception of the system is blurry: be it Apple's Siri or an online chatbot, to us it is a device, or even a tool, that, unlike traditionally perceived tools, interacts with us in a human-like way through voice, text, or other significant medium. This blur often masks the reality that we consciously need to bring forth in our minds – the device we are accessing this "intelligence" through, its algorithms, data, and the cloud system supporting all of this need to be *physically* created and maintained in the form of wires, optic cables, metal cases, plastic wraps and such for the whole to function. Much like people can't exist only as a body or mind in separation, artificial intelligence cannot exist without its specific form of embodiment. But why is the physicality of artificial intelligence something to constantly keep in our minds if it is only marginally similar to our own? The simple answer is "cost". The cost of the body of artificial intelligence as well as its creation process that we invented and utilize, is far more draining for our environment than most of our latest creations are. "A full accounting for these costs is almost impossible, but it is increasingly important that we grasp the scale and scope if we are to understand and govern the technical infrastructures that thread through our lives" (Crawford & Joler, 2018, 3).

In *Atlas of AI* (2021), Kate Crawford identifies many of the stages that physically constitute the entire domain of "AI". They form a global chain that we can follow in order to understand how artificial intelligence is created, how much its creation requires, who pays the price, and whose purposes it serves. First, we look at the excavation of rare earth minerals, extraction of elements, and other materials that sustain the hardware of artificial intelligence. One of the most important materials that is currently being excavated throughout the entire world is lithium, which is the founding element required for the manufacture of rechargeable batteries found in the vast majority of everyday gadgets, such as smartphones or computers, where artificial intelligence is accessed by its end consumers. The other ones are nickel and copper, which are being used for wires, motherboards, or as enhancements to lithium-ion batteries so that they can last longer or function more efficiently. There are also other high-demand minerals such as dysprosium, neodymium, germanium, cobalt, and dozens more. Then, we have the seventeen rare earth elements[6] excavated for processing, combining, and embedding into the gadgets and other hardware. These rare earth elements are yet to be

[6] Lanthanum, cerium, praseodymium, neodymium, promethium, samarium, europium, gadolinium, terbium, dysprosium, holmium, terbium, thulium, ytterbium, lutetium, scandium, and yttrium.

matched or substituted as the basis for our technological hardware by any other metals we know of, making them irreplaceable to date (Ibidem, 25-33). We can easily see how this founding layer of artificial intelligence is the very definition of the word *physical*, as it uses the very elements our Earth is constituted by, relentlessly tearing them from its body until there is hardly anything left to mine.

Just as the Earth consists of minerals and elements, artificial intelligence couldn't exist without extensive amounts of human labor. Regarding the exploitation of human labor in the mining industry, it has been extensively studied and criticized by various authors including John R. Wood and his work *The Miners of Windber* (1996), Gavin Hilson's publication *Small-Scale Mining, Poverty and Economic Development in Sub-Saharan Africa* (2009), Marco Aurélio da Costa's *The Gold Mining Industry in Brazil* (2022), and many more. These studies, besides many other important topics, give an overview of how human labor collectively constitutes and keeps running industries and businesses without being given any credit or adequate compensation. The same can be said about the much less visible and talked-about human labor that both Kate Crawford (2021) and Sarah Fox (2023) discuss in their works. To an untrained eye, artificial intelligence systems are seemingly automated as they give us an answer to every piece of information we demand and perform tasks in a split second. This façade however, much like that of cloud systems, is artificial and pushed forward by the tech industry and its companies as a form of appeal towards potential investors and consumers. The unseen work of many online piece workers who label, correct, tag, rate, and review these so-called automated systems and their content is what enables them to function in a way that resembles human intelligence. Without this ghostwork and millions of people doing it, many of the automated systems we use would crumble or spit out nonsensical answers (Ibidem, 53-87, especially 81). Although this work is perceived as an action, therefore bearing no physical form, the people who perform this work and actions that are gathered under this term do. In fact, their exhaustion, salaries, and working conditions are all material as well. In this instance, we can see how the physical constitutes and shapes what we perceive as disembodied, automated, and intelligent.

In the first chapter, I have briefly spoken about data processing and storage of the software behind artificial intelligence systems. But what exactly are data, and can there be other layers to their physicality other than the fact that they need to be processed and stored on physical devices that consist of physical materials and human labor, as we have observed? The word data comes from Latin *datum* which, according to my shallow knowledge of this language, roughly translates to either given or granted. According to the Oxford English Dictionary (2023), the first ever recorded use of the word data comes from the first half of the 17th century, greatly predating computational science. There are many possible definitions and connotations, according to the respective field one would want to use the word in with adjectives such as numerical information, something obtained by scientific work, something used as the basis for reference, an analysis, a calculation, something known, a fact, an assumption for further inferences, anything presented to the mind or senses, or a basis for measurement. In computing specifically, data can be understood as “quantities, characters, or symbols on which a computer performs operations, and which may be stored or transmitted in the form of electrical signals and held on recording media” (Ibidem); the exact data we are talking about concerning artificial intelligence are all various forms of image, text, audio, and video. An article from Yann LeCun (2015) summarizes the functional role data have in training artificial intelligence systems, stating that while using any training strategy, for example, machine learning, deep learning, speech, or object recognition that were already mentioned in the first chapter, large amounts of data are a founding stone because they constitute the digital material needed for the system to arrive at its conclusions through trial and error. The training process itself depends on the various algorithmic approaches, hardware, and actual coding. It generally follows this process: the system in question is intended to recognize a certain input in the form of images, videos, the likes, or a combination of these. Then, a method for recognizing these inputs is coded and large amounts of data are fed to the system, which then flags the output with a label of their name. The program then receives human feedback on the correctness of labeled inputs and the process continues once again until a suitable precision in the correctness of answers is achieved. This recognition is based on a multi-level representation of an input, which consists for example not only of various colors, or positions of objects, but saturation, and pixel positions as well. With enough representations, very complex functions can be learned and applied to different learning techniques or data sets (Ibidem, 436-442). As I have just very simply shown, this is how large data sets and programs trained on them constitute the digital foundation of artificial intelligence. At first, this training method and approach seems harmless, but as Crawford

(2021) illustrates, even data emerge from a physical world, namely in two instances – the data originate in our world and the whole process requires gigantic storage spaces. All of the pictures, voice recordings, text, and videos originate from our real world and the people who create them. The act of capturing a real-world situation on a gadget or camera is itself the very turning point when a real-life object or scenario, say a view of an apple, becomes digitalized through photography stored on a device, and therefore the line between physical and metaphysical, or non-tangible, becomes blurry and our perception of a problem often disappears completely; the problem itself though, does not. No computing data used to feed machine learning-based systems could exist without people and physical objects. But all of this data and even the cloud that acts like a metaphysical and infinite storage for them need to be housed somewhere because they cannot simply stay on the gadgets we use daily due to limitations in the technological components that would make them obsolete in size and unusable as portable devices - this task falls on large-scale data centers. They are physical buildings or manmade structures composed of a huge number of servers with technological components able to store and access data from one, or often multiple social platforms and internet websites. Similarly to the hardware of artificial intelligence, data centers also need to be constructed and built using physical work, materials, minerals, energy, and other sources (Ibidem, 89-123).

As the last example, we are going to look at the supply chain of these minerals and elements that constitute the physical body of artificial intelligence. We commonly understand supply chains as systems consisting of organizations, people, and facilities that distribute materials or services from a company to their customers. As Baldwin (2012) points out in his work on global supply chains, these emerged and gradually intensified in our modern global civilization because of the many benefits that accompany this business model. Among others, benefits include an ability to reduce the price of product manufacturing and cost-saving through business compelling in foreign markets, increased profits via strong and stable supply chains, an uptake in customer satisfaction through vast product choices and services, an enhanced cash flow through better business functions and profitability, and the improvement in the financial position of companies. Crawford (2021) identifies a cargo container used for shipping as the single item that enabled the boom of global logistics and the transformation of Earth into a global factory and consumption space. The cargo containers themselves are built from basic earth elements and carry minerals, fuel, hardware, or consumer artificial intelligence devices around the planet, without stopping for a single day. One of the cheapest options for global logistics is a cargo boat, each carrying up to 250 million tons of cargo and housing many manual laborers (Ibidem, 46-48).

There are surely many more instances we could observe and further connect to the fact that artificial intelligence is a phenomenon that is physically embedded in our world, even though it may not seem so at first glance. The peculiar examples of excavation, human labor, data, and the logistical layer of the artificial intelligence industry are somewhat special because, through their physical realities in our world, they shape and influence it, as they shape us, via their negative environmental and ethical impact and costs.

Now that we know that artificial intelligence is, at its core, deeply physically rooted, we can show how its physicality in turn transforms into the negative environmental and ethical impact it has.

4 The hidden repercussions of the AI industry

We have learned artificial intelligence to be deeply rooted in our world - I will now elaborate on how this physical reality causes many environmental and ethical problems. For this task, we will look back at the instances we identified in the previous chapter and try to see beyond the horizon of their physicality, and towards the pervasive impact they have on people and communities, not only the Earth and the environment itself.

Mining activities, as we have observed, allow the existence of hardware artificial intelligence runs on. The trajectory of this hardware, as well as in any other product, can be identified in a life cycle assessment, a “tool to assess the environmental impacts and resources used throughout a product's life cycle, i.e., from raw material acquisition, via production and use phases, to waste management” (Finnveden, 2009,). In the following paragraphs, we will trace the few first steps of the life cycle assessment of artificial intelligence hardware, starting with the excavation of minerals from which it is made. As Agata Fugiel (2017) points out, excavation comprises the extraction of solid, liquid, or gas-occurring minerals, as well as mining support (Ibidem, 159). Terry Norgate and Nawshad Haque (2010) follow up these various mining and mineral processing stages more closely – they include ground drilling for exploration, blasting as an aid for the removal of excessive material, ventilating contained or stale air in mines and shafts, dewatering the mines, loading, and haulage of the material on and off-site, auxiliary equipment used for road building and maintenance in mine sites, crushing and grinding the excavated material and the final separation of wanted valuable substances from other debris (Ibidem, 266-270). All of these methods follow different procedures and usages, yet they share a common feature of extensively polluting the environment through the usage of fuel, explosives, electricity, and water (Ibidem, Table 2). Kate Crawford (2021) also talks about the extreme amounts of waste toxins from the excavated earth, toxic waste powder from ore processing, and chemicals that pollute land and sea (Ibidem, 53-89). The resulting types of environmental damage vary greatly. Isioma Aigbedion and Samuel Iyayi (2007) identify the following types: Aerial, land, and aquatic pollution in the form of oil spillages, gas flares, dust smog, and others. Alongside comes the disturbance of flora, which is often the first casualty due to deforestation and terraforming activities on mining sites, and fauna that either feeds on these plants or depends on them for cover; noise and environmental pollution also play a huge role in driving animals away or directly killing them. Another negative effect of mining is the destruction of natural landscapes and the creation of barren lands with waste that cannot be easily disposed of. These types of activities could trigger geological hazards in the areas, such as landslides, subsidence, flooding, erosion, and tremors together with their secondary effects. On top of that, in some mining sites, there exists a chance of radiation hazard due to emissions from radioactive minerals, which causes health hazards for all organisms; some of those minerals are a by-product of tin or lithium mining and due to lack of marketability, they are often abandoned in these sites (Ibidem, 35-37). Eisler (2003) also discusses the effects of mining sites on the areas around them, as pollutants such as mercury can poison drinking water, soil, or edible creatures such as fish (Ibidem, 335); according to Crawford (2021), these effects of mining practices greatly enhance the loss and displacement of communities living near those excavation sites (Ibidem, 53-89). Both Mensah (2015) and Aigbedion (2007) also discuss how small illegal mines in developing countries often significantly disturb local ecosystems, because they are unregulated, and create dangerous conditions as well as substantial exploitation of labor workers on these mining sites.

We should not forget the human costs. Human labor behind the artificial intelligence industry is vast and many workers are unseen or hidden on purpose by tech companies to present artificial intelligence as clean, autonomous, and intelligent. Workers in the mining sector face many physical as well as psychological health-related problems such as dengue fever[7], silicosis[8], tuberculosis, cancer, radiation poisoning, rabies from bats and other animals, malaria, hearing loss, bacterial diseases, extreme physical and social depravation, traumatic injuries or death as Eisler (2003) shows in his article. The main workforce in artificial intelligence systems, alongside the engineers and programmers, are online piece workers. Piece work is a work for which the laborer is paid based on the amount produced, as opposed to the now traditional fixed wage (OED, 2023). Lacey (2007) in his contribution to the Occupational Medicine journal describes piecework as existing in “low-paid manual industries” (Ibidem, 430) and states that “although it may appear to be an outdated form of employment in 21st-century Western cultures, piecework is increasingly applied to

[7] Mosquito-borne infectious disease.
[8] Fibrosis of lungs.

home/based workers […] and seasonal and migrant agricultural workers" (Ibidem). Kate Crawford (2021) then shows the exploitative work in the artificial intelligence industry that scholars such as Mary Gray, Sid Suri, or Lilly Irani call *ghost work* or *human-fueled automation*. This type of piece work consists of repetitive and digitally performed labor, for example, labeling training data, reviewing unfitting content, or categorizing image training sets; it underlines the artificial intelligence systems, further fueling the miraculous image of artificial intelligence without anyone recognizing value contribution of digital laborers from all around the world that make these systems operational. On top of being overlooked, this work is rarely sufficiently compensated. Artificial intelligence systems build their functionality precisely on this kind of work that has to be performed manually through task platforms such as Clockworker, Figure Eight, or Amazon Mechanical Turk. Falsification of intelligence goes as far as hiring workers to directly pretend to be an artificial intelligence system through long shifts and constant checking and rewriting of those systems or their answers – all for a perfect aspiring image that can be presented to potential investors and media. We, as mere end-users of these systems, also perform these micro-tasks through tools like Google's reCAPTCHA, as we help artificial intelligence systems recognize objects in images by clicking on them (Ibidem, 63-68). This only goes to show how the artificial intelligence systems of today are heavily reliant on the exploitation of human labor, making them neither intelligent (as we have seen) nor artificial – in fact, they are human-dependent and non-rational. From an ethical standpoint, it is important to realize that the underpaid and exploited labor of "mine workers, repetitive labor on the assembly line, the cybernetic labor in the cognitive sweatshops of outsourced programmers, the poorly paid crowdsourced labor of Mechanical Turk workers, and the unpaid immaterial work of everyday users" (Ibidem, 69) is what drives the systems in which we continuously put so much funding, effort, trust, and decision-making power.

Coming back to the issue of data, there are two prime instances of their negative impact on our world: the sociological one, which involves people, and the environmental one, concerning our planet; in the previous chapter, we have tied data to physicality in many crucial instances. Let's now explore the negative environmental impact data have. As we have seen, data themselves need to be stored and carried on hardware, that is made of elements and minerals, and these need to be excavated and processed first, which we have shown to have devastating effects on our environment, as mentioned above. This downside heavily involves hardware, more specifically data centers, that contribute to pollution in another unexpected way. As Van Le (2024) points out, data centers consume vast amounts of energy, in the form of electricity, and water to power and cool down the servers they are housing; it should be noted that the grid electricity used to power these infrastructures comes from coal, gas, nuclear, and renewable energy. The main reason for such high consumption of energy is the ever-running cooling systems that maintain a stable temperature for proper functioning and the prevention of overheat of the informational technologies and equipment which could cause server malfunctions, damage, or slowing down the computing performance of servers and other related hardware. Cooling systems can take the form of chillers, cooling towers, cooling coils, or water pumps (Ibidem, 1-7). As Albert Greenberg (2008) observes, data centers are mainly automated as the average ratio of staff members per server is 1:1000; data centers house about 50000 servers on average. The main reason for this automation is explained by the fact that human performance causes large fractions of error (Ibidem,1-6), and as Van Le (2024) pointed out, the cooling systems are automated due to the ongoing regulation and temperature checking, which would require too much work if done manually (Ibidem, 1-4). According to the United States data center energy usage report (2016), the average total United States data center electricity consumption of all space types[9] was about 67 billion kWh per year (Ibidem, Figure 22). The total[10] electricity consumption in the United States in the year 2014, which includes servers, storage, network, and infrastructure, averaged nearly 70 billion kWh per year (Ibidem, Figure 21.). Avgerinou (2017) then compares the average consumption estimations and projections from European, American, and global data centers. The European consumption in 2010 was 72.5[11] TWh and the estimated amount in 2020 is at 104 TWh[12]. The United States' consumption in 2013 was 91 TWh[13], and the estimated consumption in 2020 was 140 TWh[14]. Global consumption in 2012 reached 269 TWh[15] of electricity (Ibidem, Table

[9] Concerning datacentres, space type refers to the allocated IT equipment and infrastructure, i.e. *white space* is associated with the active housing and operation of IT equipment on site of the datacentre facility, etc. (DGTL Infra, 2024).
[10] Total consumption means the total consumption of datacentres including external power-dependent sources like off-site cooling, bringing water for on-site cooling etc.
[11] 72.5 billion kWh.
[12] 104 billion kWh.
[13] 91 billion kWh.
[14] 140 billion kWh.
[15] 269 billion kWh.

1). As for carbon emission created by water consumption used for on-site cooling systems and off-site electricity generation, Crawford (2021) describes it as "[…] full of battles and secret deals, and as with computation, the deals made over water are kept close", describing how estimated consumptions made by companies that own data centers are either not shared, redacted from public records or highly inaccurate (Ibidem, 41-46). Pengfei (2023) argues that the water usage by data centers from tech giants like Microsoft, Google, and Meta reached 2.2 billion cubic meters in the year 2022, according to an estimate – that would equal the total annual water withdrawal of Denmark times two; 1.5 billion cubic meters of that water is withdrawn in the United States. Artificial intelligence undoubtedly represents one of the most prominent and fastest-growing industries, therefore the forecast estimates of operational water withdrawal for the year 2027 might reach anywhere from 4.2 to 6.6 billion cubic meters globally. For context, OpenAI's GPT-3 consumed 700,000 liters of water during its training phase. Since its deployment, it requires "[…] a 500ml bottle of water for roughly 10-50 responses, depending on when and where it is deployed" (Ibidem, 1-3).

As I have suggested, the semantically murky term *data* has, much like artificial intelligence itself, more layers to it than just environmental impact. We now turn to address its second instance hinted at earlier, that being the impact it has on us, people. Crawford (2021) tackles this problem quite remarkably – according to her research, the data scraped from the internet, as well as individuals, sold by companies, or acquired by other means, could be considered *stolen*. Not only are various images, text messages, or sound recordings taken from users without their consent, but they are also stripped of all context and details that make up our understanding of the world and instead are fed to commercial and academic models only as *some* data. To make things worse, people depicted or by other means involved in photos and other forms of data have no say in how these are used and aren't even aware that their digital imprint has become food for artificial intelligence testbeds (Ibidem, 89-96). This brings forth multiple problems, such as the unethical and illegal use of personal data, non-consensual data mining, the conscious loss of the semantic and context value of data, and the potential of replicating bias in the consecutive labeling of the now-training data in large data sets. Now, let us turn to take a closer look at these issues.

Crawford (2021) talks about how scarce and rare consent has become concerning the usage of data concerning us and its extraction from the internet. According to her, it "[…]

changed everything; it came to be seen in the AI research field as something akin to natural resource, there for the taking […] anything and everything online was permitted to become a training set for AI" (Ibidem, 106). One of the biggest turning points was social media platforms, where suddenly millions of photographs or videos, labeled with names or locations, were available for the taking and then used by big tech companies to feed artificial intelligence algorithms or sold elsewhere to create large datasets. Not very long after this data boom even more problematic extraction methods began to emerge, including instances where a professor secretly installed a camera on campus walkaway to capture photos and videos of people for the sake of training his facial recognition system, the Duke University non-consensually capturing footage of their students with financing from U.S. Army Research Office and then publishing the results on the internet, or researchers commandeering a camera to capture everyday life in a cofee without anyone's consent. There are a staggering amount of these cases, and they are by no means isolated or rare (Ibidem, 105-111). To highlight the impact these abstract consent and data concepts have on our world, let us consider other such events, this time on a much larger scale. In recent years, I have observed many of them, including the Clearview AI Facial Recognition violation of privacy laws and data protection regulations in 2020, where its artificial intelligence model scraped billions of images from the internet without users' consent in order to build a facial recognition database, which was later sold to law enforcement agencies and private companies under undisclosed purposes for the use of scraped data. Earlier that decade, in 2018, Facebook AI and Cambridge Analytica joined forces to create psychological profiles for political advertising, thanks to the mined data of Facebook users, to influence voter behavior during the previous U.S. Presidential elections and Brexit referendum. During the same year, IBM Watson faced a backlash as it was shown that their artificial intelligence system for medical advice in oncological treatment was not properly trained on real patient data and therefore recommended potentially harmful medical advice. Then, in 2017, Google's AI subsidiary, DeepMind, accessed around 1.6 million patient records from the National Health Service in the UK without their consent to develop a detection app preventing further acute kidney injuries. Hutton and Henderson (2017) argue that companies often collect, mine, distribute, and process large data volumes without the knowledge or consent of individuals to which the data pertain. The authors make a strong point about the fact that existing privacy notices and user agreements are nothing but incomprehensible gibberish to ordinary people and allow the data collectors high flexibility in what practices they can use, and how the data is stored or further distributed. Individual human subjects can neither control these invisible data flows, terms of their use, nor correct any potential misinformation in them. Even academic researchers often make use of these datasets, obtaining approval from their respective institutional review board that relies on post-war ethical protocols unsuitable for the era of the Internet. According to the authors, this makes the existing consent instruments insufficient or even unnecessary, since they are overlooked and often not regulatory enough (Ibidem, 147-154).

Crawford (2021) stresses the conscious loss of data's semantic and context value. On this account. She gives an example of a Multiple Encounter Dataset that tracks criminals or arrested people through biometrics and photos, which are then turned into data for further processing. She makes a strong point that people presented in the databases' mugshots are shown as simple data points without further explanation of their arrest, their names, or any context to help us understand if they were charged, arrested, or imprisoned. People in this dataset are turned into data used to improve artificial intelligence tools, such as the face or affect recognition. This poses another problem alongside the loss of consent, as the data scraped from the internet or such databases are stripped of all context, meaning, and history – the trained systems can then tell the difference between an apple and a human being but cannot explain why the individual was inside a building, say a police station (Ibidem, 91-95). Consequently, "any individual image could easily be substituted for another and the system[16] would work the same" (Ibidem, 94). The mindset behind such practices ties directly to an earlier mentioned view on artificial intelligence and to the claim that it is not at all intelligent or self-governing, as it needs human intervention and labeling to blindly replicate this context. This situation is not fundamentally different from the one Georg G. Iggers wrote about in his book *Historiography in the Twentieth Century* (2002). He talks about a certain linguistical turn that, in the eyes of some postmodern historians and critics, needed a rational and critical revision (or re-evaluation). The revision concerned the relationship of historians to historical sources, culture, and meanings, as it was often too influenced or biased by the author's beliefs and cultural roots and therefore explained in ways that might lead the actual meaning astray

[16] *The system* is referring to any artificial intelligence system operating based on this data, for example the mentioned facial recognition field.

(Ibidem, 111-124), potentially causing historical-based discrimination towards some groups or being used as a tool of power is a similar manner – with the revision of this turnover, we have learned to look at such thing more objectively.

As we have already observed, and as Crawford (2021) demonstrates, data sets are the key to training any artificial intelligence and machine learning systems, i.e. facial or object recognition systems. The most important thing is what the training data consists of – Crawford uses an example with apples, saying that if we were to include only images of red apples, not green ones, the learning system would never flag a green apple as an apple, as it only operates with the data available to it, not following a logical premise or generalization. Then, through competitions and various testing methods, including performance and percentage accuracy measurements, training data sets are established; what follows is their adaptation, further building upon them, and their eventual user expansion – it creates a rather layered image of newer forms of these datasets being constantly added, while the older ones at the core stay roughly the same (Ibidem, 96-98). One of the most notable examples is ImageNet which, according to Deng et al (2009), is a "large-scale ontology of images" built upon the structure of an older semantic database called WordNet that aimed to contain roughly "50 million[17] cleanly labeled full resolution images (500-1000 per synset[18])" that it scraped from the Internet. It is divided into 12 subtrees and more than a thousand image classes, all labeled according to the thing they represent (Ibidem, 248-251). So how exactly does bias express itself in artificial intelligence systems, or rather in the data sets they are trained upon? Prejudices can take on many forms, such as the shape-texture bias that Robert Geirhos et al (2018) discuss. They show how convolutional neural networks (CNNs) trained on the ImageNet dataset are "strongly biased towards recognizing textures rather than shapes, which is in stark contrast to human behavioral evidence and reveals fundamentally different classification strategies" (Ibidem, 1). They continue to show how a texture-shape conflict can greatly affect these recognition systems, by putting an elephant skin texture cues on the shape of a lying cat – to a human eye, it is a cat, yet more than half of selected CNN responses say

[17] According to other up-to-date articles, this number is around 14 million labelled images.
[18] Sometimes also called a synonym ring, is a group of words interchangeable with one another (semantically equivalent), depending on the context they are used in. For simplicity's sake, the synset for "sad" (interchangeable with the words unhappy, regretful, down, etc.) on average contains 750 images tied to the concepts of sadness, i.e. sad people, animals, etc.

“Indian elephant” as an answer to said image (Ibidem, 1-6). Another aspect of this many-sided dice, according to Logan Engstrom et al. (2020), is the statistical bias they study on the replication of ImageNet, ImageNet-v2. According to this study, dataset replication[19] in machine learning, though not so common, took inspiration from experimental studies of natural sciences. Recent dataset replication studies found an alarming drop in the rate of replicated model performance and accuracies, averagely amounting up to 11% on the lower end, and 14% on the higher end. Engstrom et al. identified a possible mechanism that causes such a drop: noisy data readings. In short, the replicated dataset needs to be revised for any new data of meaningless information value and statistical matching of the data it already contains; an example might involve incorrectly labeled data caused by the replication process pipeline. This, in turn, mitigates most of the performance and accuracy drop to around 3% (Ibidem, 1-4). There are of course many more kinds of bias in the artificial intelligence field, discussed in the article of Judy W. Gichoya and others (2023). One of the biggest problems identified in this work is the “[…] unified definition of bias” and “focus on a statistical definition of bias and a technocentric definition […]” which leaves the recipients of this impact, individuals and communities, largely omitted. Another huge problem this article expresses is the lack of, or in the best case a very little amount, of effort put into bias monitoring after these systems are deployed for the broader public or any other end-user. Some examples of this may include bias based on race, gender, age, education, history, cognition, and inheritance from other models – all of these can then be differentiated to either human or machine biases occurring at each stage of artificial intelligence systems development (Ibidem, 1-8). It is once again vital to point out the real and physical impact of bias and prejudice that occur in artificial intelligence systems, on our world and the people living in it. Some of the most notable examples one can learn about include artificial intelligence-based algorithms more often than not discriminating against women in their automated job applications or automated discrimination against minorities in the banking sector. Then, we have the COMPAS, or Correctional Offender Management Profiling for Alternative Sanctions, management, and decision support tool used in the U.S. jurisdiction to determine a likelihood defendant becomes a recidivist. In 2016, a scandal arose as several articles and individuals demonstrated that the system clearly discriminated against black

[19] Not to be confused with dataset-building or being based on other already existing datasets.

people, and to the contrary underestimated the reoffending rate in the case of white people; there was also a sex-based bias underlying these claims (Hübner, 2021). Our next example also includes racial and ethnic bias but in the environment of healthcare algorithms. Findings suggest that algorithms used and implemented in healthcare sectors can, and often do, underestimate the needs and care given to people of color compared to white patients (Jain et al., 2023). Last, but definitely not least, we have the 2016 Microsoft chatbot incident. Upon its training on the closed dataset the chatbot named Tay was supposed to learn more through unrestricted human interaction on Twitter[20], yet it took less than 24 hours for the project to be withdrawn as it became openly racist and sexist through targeted corruption by the users of the said platform (Wolf et al., 2017).

All the examples above are underlined by the unethical and potentially illegal use of training and generated data. Kate Crawford (2021) calls it “ethics at arm’s length”, drawing inspiration from Joseph Weizenbaum, an artificial intelligence scientist and critic. The thought behind the concept might seem quite simple, yet it is vital to be vary of it. To this day, artificial intelligence fields and research elude any greater review. The lack of review rests on the historical embedment of numerical sciences being looked upon as something detached from our human nature. But as we have seen, this logic cannot apply in today’s times, as individuals and groups feel the consequences of data-driven algorithms firsthand. And so, we get to the merit of ethics at arm’s length, which is the detachment of researchers' invasive extraction of data from their subjects without interacting with them, let alone realizing the ethical and other consequences of their actions because of this separation (Ibidem, 115-117). This distancing from the ethics and consequences, paired with illegal and non-consensual data mining and personal data leaks or trading orchestrated by company customers and network users (Vishnevskaya et al., 2021) created a mindset hellbent on quantity, a colonial attitude heritage, and progress at all costs that drives and governs the field of artificial intelligence.

Our last showcase of physicality entwined with hidden repercussions are the supply chains and cargo that support the colossal, global, and ever-hungry artificial intelligence factory. Globalization, capitalism, and market surplus aren’t by any means new or neglected subjects, yet they should resonate through every debate about artificial intelligence. These concepts can be applied to any aspects we have observed, be it globally obtainable data, a

[20] Nowadays known as the “X” platform owned by E. Musk.

surplus of artificial intelligence algorithms at the cost of their ethical qualities, or the fact that for late capitalism, everything, including creating artificial intelligence, is a business and another sign of successful technological progress. In light of availability, the supply chains, mainly marine shipment of cargo containers, enable this latest giant boom in the commercial artificial intelligence industry. From the shipment of rare earth minerals and elements to components, workers, technological devices that artificial intelligence runs on, and cooling and powering devices to be available almost everywhere, the global shipment industry plays a crucial and yet largely unknown role. According to the European Commission, global shipping emissions stood for 1.076 million tons of CO_2 or 2.9% of global emissions caused by any human activities; at the level of the European Union, marine shipping averages 3.5% of total CO_2 emissions (EUC, 2024). The global CO_2 emissions from logistical transport account for around 20% of global carbon dioxide emissions. That is, if we only consider CO_2 emissions from energy, the real number can come up much higher, according to Ian Tieso (2023). The overall damage, alongside emissions, is once again vast, as not only our air but also marine and land life is displaced, species go extinct, and the geological composition of land and salinity of water is changed. Overall, the logistical industry is as slippery as the mining sector in terms of pollution and environmental damage.

The complexity of what artificial intelligence is and what it consists of, as we have shown, is quite hard to accurately describe, as Tamburrini (2022) points out: "Altogether, one can hardly question the importance of estimating the AI carbon footprint and identifying its sources. However, providing realistic estimates presupposes an inventory of the wide variety of relevant factors and the development of suitable carbon footprint metrics and models" (Ibidem, Section 2). Said "relevant factors" include the move towards transparency from tech companies that Crawford (2021) talks about. This can include the explicit sourcing of the mined materials, shipment used, data collected, and much more. According to Cowls et al. (2023), the Information Communication Technologies sector, or ICT for short, is responsible for 1.4% of global greenhouse gas emissions and in the year 2030 can account for up to 23%; note that this is projected on a study from 2008.

Even though the polluted rivers, stripped lands, misplaced communities, an air filled with invisible dust particles, biodiversity loss, and hard work of people of those locations may seem far away from us and thus largely irrelevant, we need to acknowledge the fact that while we use any hardware to access artificial intelligence systems, we do so on behalf of our environment and individuals or groups that are being continually affected. This statement is not supposed to be a blind plea for stopping all technology production or a misguided hope for miraculous healing of the planet, but rather a desire for change in the mindset of people. I believe that by acknowledging the real costs of our gadgets, we can look at the future in different ways and perhaps start a transformation that will lead towards great environmental responsibility and reshaping of our current methods in sectors such as the technological and mining ones.

5 New ethical imperative

Throughout this thesis, we arrive at a point where we (hopefully) see artificial intelligence itself, its underlying systems, and its various shapes and shades in a different and new light. So far, thanks to the many works of cited authors, I have explored how artificial intelligence should not be taken only as the latest technological gadget, but as a strong force to be reckoned with. This force can and does negatively alter the lives of individuals and communities through its extensive and more frequent commercial use in places such as medical recommendations, hiring interviews, judicial processes, statistics-based predictions, and much more. It also influences the environment as well as many life forms through an extensive excavation of materials, processing, transportation, manufacturing of hardware it runs on, storing its content in the form of data centers, and constant maintenance. This hidden physicality of artificial intelligence, on top of being heavily labor-dependent to seemingly function as autonomous, is enormously costly. These expenses, often invisible to the naked eye of an end-user, take a toll on humans, energy, animals, and the environment alike.

Personally, I see artificial intelligence as a great and potentially revolutionary tool that can help us in combating the still worsening and time-delicate threat of climate change, whose consequences we are more than sure to reap. But when I discovered the reality of this magical intelligent and autonomously presented system, I sobered up from any premature optimism as long as its problems remain undiscussed and overlooked, oftentimes even disguised through the vision of more profitable sales. Artificial intelligence, as it is now, cannot amount to anything but a negative influence. How can a system this costly, often biased towards people, privately held, people controlling, and used for power-attaining purposes aid us in resolving a problem that requires unity, transparency, cooperation, and a common goal? In this case, the proverb “fighting fire with fire” simply cannot work.

That is, it cannot work on the presumption that artificial intelligence itself turns out to be what everybody expects of it – a human-like sentient and autonomous, or artificial general intelligence that can one day surpass us. If this does not come to be, how are we to justify enormous resources of all kinds put into it? What I am getting at is a techno-positivist argument that artificial intelligence, or rather artificial general intelligence, can reshape society, resolve climate change and overall help humans take the next evolutionary step. In my thesis, I try to be more modest asking "But how do we know that exactly?" This attitude can very well be positive, just as we have hinted upon it, but it can very well end in a disaster that we cannot yet imagine. Is the cost of sustaining these vast and complex artificial intelligence systems really worth some potential and abstract gain we can obtain from supporting them?

My new ethical imperative does not offer or state a final resolution of this matter – not only do I not have the competence to do so yet, but I also strongly doubt that an assembled team of professionals could because of the vastness and complexity I have but scraped the surface of. Instead, through my work, I propose a change of thinking and perception of artificial intelligence, accepting its physicality and environmental impact. For it to be ever of any use to us, we first need to accept that we bear the consequences of not only our posterior actions but our prior creations too – and if these creations represent such poor values and forms of use, then the first place to change is inside of our mind. The new ethical imperative towards artificial intelligence should make visible not only the environmental costs but also the often-unspoken exploitation of people participating in keeping this global mega-machine operational. The most critical point is that we all, without exceptions, should keep these costs in mind and be more careful in the ways we use artificial intelligence as well as think what we want it to become. If we wish for artificial intelligence to help us manage our biggest existential crisis yet, we need it to not continually delve deeper into it by polluting the very environment and damaging the very communities we aim to preserve and grow. This is one of the key aspects that the new ethical imperative towards technology such as artificial intelligence should take into account. To cite one of my favorite professors, an environmentalist and a co-founder of my faculty, Dr. Ivan Rynda: "We will not be redeemed by specialization, such as technologies on one hand or just moderation on the other hand (for example in the food industry or in the way people cater food) but always and only synergy" (Zdráhalová, 2024; translated by the author).

Literature

1. Aigbedion, Isioma, and Samuel E. Iyayi. "Environmental effect of mineral exploitation in Nigeria." International Journal of Physical Sciences, vol. 2, no. 2, 2007, pp. 33-38.

2. Arendt, Hannah, and Václav Němec. "Vita activa, neboli, O činném životě" [The Human Condition]. OIKOYMENH, 2009.

3. Avgerinou, Maria, Paolo Bertoldi, and Luca Castellazzi. "Trends in data centre energy consumption under the European code of conduct for data centre energy efficiency." Vol. 10, no. 10, 2017, p. 1470.

4. Baldwin, Richard E. "Global supply chains: why they emerged, why they matter, and where they are going." 2012.

5. Beik, Mildred Allen. "The miners of Windber: The struggles of new immigrants for unionization, 1890s-1930s". Penn State Press, 1996.

6. Coeckelbergh, M. "AI Ethics". MIT Press, 2020.

7. Cowls, J., et al. "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations." 2023, pp. 1-25.

8. Crawford, K. "Atlas of AI." Yale University Press, 2021.

9. Crawford, Kate, and Vladan Joler. "Anatomy of an AI System." 2018.

10. Deng, J., et al. "ImageNet: A large-scale hierarchical image database." 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi:10.1109/CVPR.2009.5206848.

11. Dietmar, Hübner. "Two kinds of discrimination in AI-based penal decision-making." Vol. 23, no. 1, 2021, pp. 4-13, doi:10.1145/3468507.3468510.

12. Engstrom, L., et al. "Identifying statistical bias in dataset replication." Proceedings of the 37th International Conference on Machine Learning, 2020, pp. 2922-2932.

13. Eisler, Ronald. "Health risks of gold miners: a synoptic review." Vol. 25, 2003, pp. 325-345.

14. Finnveden, Göran, et al. "Recent developments in life cycle assessment." Vol. 91, no. 1, 2009, pp. 1-21.

15. Flasiński, M., & Flasiński, M. (2016). "History of artificial intelligence: Introduction to artificial intelligence."

16. Fox, Sarah E., et al. "Patchwork: the hidden, human labor of AI integration within essential work." Vol. 7, CSCW1, 2023, pp. 1-20.

17. Fron, Christian, and Oliver Korn. “A short history of the perception of robots and automata from antiquity to modern times.” 2019.

18. Fugiel, Agata, et al. “Environmental impact and damage categories caused by air pollution emissions from mining and quarrying sectors of European countries.” Vol. 143, 2017, pp. 159-168.

19. Gichoya, J. W., et al. “AI pitfalls and what not to do: mitigating bias in AI.” Vol. 96, no. 1150, 2023, p. 20230023.

20. Geirhos, Robert, et al. “ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness.” arXiv preprint arXiv:1811.12231, 2018.

21. Goudie, A. S. “Human impact on the natural environment.” John Wiley & Sons, 2018.

22. Gottlieb, Robert. “Reconstructing environmentalism: Complex movements, diverse roots.” Vol. 17, no. 4, 1993, pp. 1-19.

23. Greenberg, Albert, et al. “The cost of a cloud: research problems in data center networks.” Vol. 39, no. 1, 2008, pp. 68-73.

24. Hilson, Gavin. “Small-scale mining, poverty and economic development in sub-Saharan Africa: An overview.” Vol. 34, no. 1-2, 2009, pp. 1-5.

25. Hutton, L., Henderson, T. “Beyond the EULA: Improving consent for data mining.” Vol. 32, Springer, Cham, 2017, pp. 7-27.

26. IGGERS, G. “Dějepisectví ve 20. století. Od vědecké objektivity k postmoderní výzvě” [Historiography in the Twentieth Century]. Praha: NLN, 2002.

27. Jain, A., et al. “Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms.” Vol. 4, no. 6, 2023, e231197.

28. Kolbert, E. “The Sixth Extinction.” Holt, 2014.

29. Lacey, Rosie J., et al. “Piecework, musculoskeletal pain and the impact of workplace psychosocial factors.” Vol. 57, no. 6, 2007, pp. 430-437.

30. LeCun, Yann, et al. “Deep learning.” Vol. 521, no. 7553, 2015, pp. 436-444.

31. Li, Pengfei, et al. “Making AI less ‘thirsty’: Uncovering and addressing the secret water footprint of AI models.” arXiv preprint arXiv:2304.03271, 2023.

32. Mensah, Albert K., et al. “Environmental impacts of mining: a study of mining communities in Ghana.” Vol. 3, no. 3, 2015, pp. 81-94.

33. Moore, R. R. “Moore's law: past, present and future.” Vol. 34, no. 6, 1997, pp. 52-59.

34. Muthukrishnan, N., et al. (2020). “Brief history of artificial intelligence.” Vol. 30, no. 4, 2020, pp. 393-399.

35. Nadakavukaren, A., & Caravanos, J. “Our global environment: A health perspective”. Waveland Press, 2020.

36. Norgate, Terry, and Nawshad Haque. “Energy and greenhouse gas impacts of mining and mineral processing operations.” Vol. 18, no. 3, 2010, pp. 266-274.

37. Oxford English Dictionary, s.v. “automata.” Oxford University Press, 2023.

38. Oxford English Dictionary, s.v. “datum (n.).” Oxford University Press, 2023.

39. Oxford English Dictionary, s.v. “piecework (n.).” Oxford University Press, 2023.

40. R. R. Schaller, “Moore's law: past, present and future.” Vol. 34, no. 6, 1997, pp. 52-59.

41. Rosie J. Lacey, et al. “Piecework, musculoskeletal pain and the impact of workplace psychosocial factors.” Vol. 57, no. 6, 2007, pp. 430-437.

42. Shehabi, Arman, et al. “United States data center energy usage report.” 2016.

43. Solomon, S., et al. “IPCC Fourth Assessment Report (AR4).” Vol. 374, 2007.

44. Strubell, Emma, et al. “Energy and policy considerations for deep learning in NLP.” arXiv preprint arXiv:1906.02243, 2019.

45. Tamburrini, Guglielmo. “The AI carbon footprint and responsibilities of AI scientists.” Vol. 7, no. 1, 2022, p. 4.

46. Van Le, Duc, et al. “Impacts of increasing temperature and relative humidity in air-cooled tropical data centers.” 2024.

47. Zdráhalová, Helena. “Rynda: Pro udržitelný rozvoj je klíčový růst chudých zemí.” Forum. Univerzita Karlova. 2024