

BACHELOR THESIS
ARTIFICIAL INTELLIGENCE

Radboud University



**Artificial Intelligence As the Solution to All Our
Problems: Why It Is a Paradigm of
Dehumanisation Instead**

Author:
Luc van der Gun
S4807987

Supervisor:
dr. Olivia Guest
Department of AI
olivia.guest@donders.ru.nl

Second reader:
dr. Annelies Kleinherenbrink
Department of AI
annelies.kleinherenbrink@ru.nl



January 27, 2023

Abstract

It goes without saying that Artificial Intelligence (AI) is at the pinnacle of our technological advancements. We have created intelligent systems that already vastly outperform certain human capacities, and it is clear that this will only increase as time passes. In many ways this is very promising, but the forms that technology and AI take in our society has also sparked many concerns of dehumanisation. On the one hand, attention is drawn towards the power play that is underlying the use of technology, such that the dehumanisation would be gone if only we rule out these hidden intentions or mistakes in the applications. Yet, such a view glosses over the fact that dehumanisation might also occur when a seemingly positive goal is perfectly reached, a phenomenon that I coin “nonintentional dehumanisation”. According to the philosopher Martin Heidegger, the essence of technology is “enframing”: a paradigm of efficiency and resourcification that enables a readily acceptance of technology, even though it is inherently nonneutral. This thesis argues how this “enframing” by AI leads to intrinsic nonintentional dehumanisation. Furthermore, by drawing from Haslam’s “mechanical dehumanisation” and Borgmann’s “device paradigm” this nonintentional form of dehumanisation is elaborated to take shape in two forms: the denial and the deprivation of a certain humanness in those who are subjected to it. Finally, in order to evaluate to what extent this problem is concretely occurring, this two-fold model is used to investigate two contemporary domains of AI applications: human resource management and healthcare. The message is that if we are to combat this paradigm of dehumanisation, we need to be wary of the values that are inherent to the AI-induced changes, and the proposed identification model could prove useful when evaluating AI applications in the future.

Contents

1	Introduction	2
	On the Anthropological View of Technology	2
	Towards Technology Itself	4
	Layout of This Thesis	5
2	On the Essence of Technology and AI	6
	Technology As a Way of Revealing	6
	The Challenging by Modern Technology	7
	Enframing As the Essence of Technology	9
	AI and Enframing	9
	The Aggravation by AI	10
3	The Dehumanising Character Underlying AI	12
	Haslam’s Mechanical Dehumanisation	13
	Dehumanisation As Beyond Denial	14
	Borgmann’s Device Paradigm	15
	A Two-Fold Model of Nonintentional Dehumanisation	17
4	Investigating Contemporary AI Applications	19
	AI in Human Resource Management	20
	Modifying and Replacing Human Jobs	21
	Algorithmic Management	22
	AI in Healthcare	24
	Algorithmic Healthcare Governance	25
	Replacement of Human Caretakers	26
5	Discussion	28
	Overview	28
	Limitations	29
	Future Implications	30
	On the Ambiguity of Human Control	30
	The Way Forward	31
	Conclusion	32
	Acknowledgements	33
	References	34

Chapter 1

Introduction

When turning a listening ear towards the oft-discussed societal implications of Artificial Intelligence (AI), one quickly notices that two opposing myths are crystallised at the very centre of the debate: either it will be against us, or it will come as a rescuer in dire need (Raisch and Krakowski, 2018). The former is argued to occur because it may perform its tasks unrestrained and at the expense of us humans – such as the famous "paperclip maximiser" (Murphy, 2018) –, because it may be used to oppress people (Kerr, 2021), or perhaps because it actually turns evil (Khole and Hande, 2018). Oftentimes the term “dehumanisation” reveals itself in order to give words to our intuitions regarding the matter, but in general it is a label that has different meanings in different contexts, thus lacking any true coherence (Haslam, 2006). The latter, more Utopian representation of AI, attributes to it the role of a saviour and perceives and prospects that will free us from the necessity of hard and tedious work, essentially displaying it as "a silver bullet to solve complex social problems" (Leufer, 2020). According to the famous psychoanalyst Carl Jung, myths display a profound influence on the way we behave and experience our life (Segal, 1998), and as such myths can be described to be affective dispositions that we have towards something en masse, revealing to us how we view it and how we relate ourselves to it, e.g. are we more hesitant or are we willing to incorporate it in our lives? With regards to computer technology, Mosco (2004) argues too that myths are of paramount importance when discerning the place of it in society. He states that although the construction of myths is a human genius to provide meaning for the past, present and future, this ability comes at the price of having the tendency to mask any underlying influences that it has on us, and by creating potentially deceptive impressions of social cohesion, our relationship towards it becomes hidden, obfuscating the present conflicts that may actually be happening. According to Mosco, myths are thus a valuable tool to see what they reveal (desires) and what they conceal (concrete issues), and our knowledge of them should be used to loosen their strong hand.

On the Anthropological View of Technology

This critical attitude towards the myths surrounding technology has definitely received attention throughout the years. Pfaffenberger (1988) for example also recognises the fact that our mythical

conception of our power over technology might be dangerous, when he states: “The culturally-supplied notion of ‘technology’ carries with it two tacit meanings, two implicit and mythic views of the world in relation to technology, that profoundly affect how we understand technology and how we view its relationship to our lives.” (p. 237) The first variant is identified as “technological somnambulism”: a view where impact of technology is merely that what we do with it. It is a position that holds that it is simply a tool that we can use whenever desired, and put down otherwise, thus disregarding the need for reflection on its usage, as technology is quite plainly perceived to be harmless in itself. This is a point that was publicised by Davis in an article in *Science* called *Technology as a Deterrent to Dehumanization*, when she wrote: “Is technology the real culprit in making the automobile workers, the students, and the supermarket shoppers feel dehumanized? Emphatically no! Technology is neutral. It does not come with a label specifying its end product or intended use.” (Davis, 1974, p. 1). The neutral identity of technology is a given for the view of technological somnambulism. The second variant is identified as “technological determinism”, that holds that technology is inherently autonomous and that it is governing our lives. It views technology as being pushed forth by science, which in turn is progressed through advanced technology, and this back-and-forth influence is taken to be happening in a perpetual circle, such that it has achieved the status of full uncontrollability. For Pfaffenberger, both views on the use of technology are clearly misleading, in that they gloss over the social relations in which technology is embedded. He views both mythic versions as forms of “Marxist fetishism” – which entails that the charm of technology is hiding the underlying social dimensions from view –, stating that through them technology is seen as mere “relations among things”. Instead, he opts for an anthropological view that identifies technology as being “relations among people”: viewing technology not as an “ideology of objects”, but as a “humanised nature”, i.e. as being fundamentally a social phenomenon. Reincorporating this saviour-role that is attributed to AI, as discussed at the beginning, Munn (2022) elaborates a very similar conception of technology as he writes against the “myth of automation”, which is a fantasy of full automation everywhere and automating everyone. He elaborates that the promise of full automation assumes a “unified vision”; supposedly it is has an effect on the universal man, which is dangerous because in actuality it gives rise to less unification, more unequal power relations and more racism instead, as he states: “In claiming to be global, automation overlooks its context. . . . And in claiming to be inevitable, automation conceals its interests.” (p. 11) Thus, the failure of the promise of AI of bringing a unifying automation is connected with implicit hidden political agendas, just like Pfaffenberger. Automation is found to reflect the distinct cultural values that presuppose it, such that its general concept is associated with the historical, social and racial dynamics at work. The automating potential of AI – the glamorous promise of a world where everything is taken care of – is taken to be fundamentally shaped by underlying favouritism and ill intentions, and as such, dehumanisation is effectively linked with power play.

Towards Technology Itself

Both Pfaffenberger and Munn aim to bring the hidden political dimension of technology, automation and thereby AI to light. In essence, they write against Davis' statement that technology is neutral, as they take that technology always has concealed agendas and thereby hidden implications. Clearly there is some truth to their point, as for example in the Western world Big Tech companies are oppressing the uninformed users by means of technology (e.g. "surveillance capitalism" as described by Zuboff, 2019), thus evidencing for the fact that social relations are constitutional to the use of technology and some of the accompanying forms of dehumanisation. But, by focussing solely on this social dimension, they seem in turn to gloss over the very nature of technology itself, which lies beyond the social context of use. The philosopher Martin Heidegger regards technology as inherently nonneutral too - stating that this nonneutrality stems from the fact that technology is not a mere means to an end (Heidegger, 1977). Though it may appear to be complying with the anthropological view of technology presented above, it is different in one crucial aspect. What the anthropologists implicitly still hold, is that this new conception of technology itself is still neutral, just like the conception of technological somnambulism. By positing concealed ends that are underlying the use of technology, they still describe technology as being merely a means to an end; it is just that the ends are now hidden from the general public. As Heidegger would say, this view of technology is correct, but that does not necessarily mean that it is the true. Proclaiming the influence of covert political intentions is a legitimate way to criticise the general acceptance of technology, but it is at least equally necessary to investigate the dehumanisation arising due to the very nature of technology itself.

Heidegger argues that the essence of technology is "enframing": a paradigm of resourcification and efficiency that inherently puts forward a reducing view of the world and ourselves (Heidegger, 1977). In a way, this critical evaluation strives to go beyond agendas, thus arriving at a world of nonintentions. It shifts the focus away from technology as being embodied malicious motives, to see instead how problems arise due to an unawareness of all humans. In its core it is not about the oppression of the few over the many, but rather it investigates how we are all affected by the promise or myth of AI of solving all our problems, in that we ourselves are "doing" this enframing too. As such, the nonintentional has consequences; a category of effects that are prevalent exactly when the goal of an AI application is perfectly reached and its method perfectly executed. Thus statistical bias falls beyond this category, as it is either an inaccuracy with regards to the goal, or if it is in fact a desired property of the system, one should again take a look at technology through a more anthropological lens. It is about the ambiguous and difficult relation between humans and technology: one could say that we are in control of it as it has no impulse to drive forward without us, yet we are also affected by it in ways that we do not immediately foresee. We should therefore look at it more in terms of "relations among people and things" rather than in terms of "relations among people" as Pfaffenberger suggests. Thus, nonintentional consequences ultimately arise due to the specification of undesirable goals that we wish the AI systems to attain, and as such the problem is inherent to the goals we put forward.

The four views on our relation to technology that were discussed up to this point are listed in table 1.1 for a clear summary of the different stances that they take. In comparison to research on the social dimension there have been relatively few investigations into how our general relationship to AI may lead to dehumanisation. Through a notion that I coin “nonintentional dehumanisation”, I aim to embed the rather sporadic, yet related, accounts of dehumanisation by AI. In order to uncover this, the values inherent to the very essence of AI should be investigated first. The question that will stand absolutely central to this thesis is: *To what extent does the readily acceptance of AI lead to nonintentional dehumanisation?* This thesis directs to be a coherent picture and a general overview of the phenomenon nonintentional dehumanisation, and aims to find out exactly in what ways and to what magnitude this predicament is the case.

Definition	View on technology	Human control	Neutral usage
Technological somnambulism	It is merely what we do with it.	✓	✓
Technological determinism	It is spiralling out of our control.	✗	✗
Anthropological view	It conducts underlying social relations.	✓	✗
Enframing view	It is a paradigm of efficiency.	✓	✗

Table 1.1

The different views on our relation to technology and their portrayal of our control and neutrality of our usage of it. Technological somnambulism separates itself from the others by envisioning technology as being fully under our control, because it accepts technology as being neutral. On the other hand technological determinism separates itself because it portrays technology as being strictly out of our control, and thus it’s usage is nonneutral. The anthropological view might seem similar to the enframing view, in that they both acknowledge human control while our usage of it is nonneutral nonetheless. However, this nonneutral usage is argued for a different reason for both: the enframing view relates it to the fact that technology itself is essentialistically nonneutral, while the anthropological view relates it to the concealed underlying social relations.

Layout of This Thesis

To lay clear the structure of the rest of this thesis I will give a short summary of what is discussed in each of the remaining chapters. In the second chapter I investigate and explicate what Heidegger means when he states that technology is nonneutral and that its essence is enframing, and I then apply this notion to AI specifically in order to see why it contributes to it. Chapter three will elaborate the two forms of nonintentional dehumanisation (the denial and deprivation of a humanness) that is inherent to this essence of technology. In the fourth chapter I will use this two-fold conception of dehumanisation to aid in making transparent and concrete to what extent this nonintentional dehumanisation is present in two central fields of contemporary AI applications: human resource management and healthcare. Finally, chapter five will conclude the apparent impact of this dehumanisation, discuss its difficulty of truly assessing it, as well as elaborate a way forward by outlining what part falls within the region of our control and to what measure this controlling is or is not desirable in the first place.

Chapter 2

On the Essence of Technology and AI

Forming the backdrop of this thesis, I will start with an analysis by perhaps the most famous philosopher of the 20th century, Martin Heidegger, in order to have a solid background against which to paint what is to come. In his work *The Question Concerning Technology* (Heidegger, 1977) he elaborates a critique on modern technology by investigating what it is in its very essence, because in his view it is only possible to have a free relation to something if one knows what that something truly is. He argues that technology is not merely neutral, that it is not simply a means to an end and that we should therefore re-evaluate our sense of being capable of remaining unaffected by it. Instead, the nature of technology (I will use “nature” and “essence” interchangeably) is argued to be revealing, or rather: “enframing”. *The Question Concerning Technology* was first published in 1954 (in German), and its ideas were conceptualised just preceding the birth of AI, still years apart from any appearance of modern AI (such as neural networks) used in our contemporary society. Heidegger thus did not write about AI, yet it is evident that AI is in every sense part of modern technology. In this chapter I will first explain why Heidegger thinks the essence of technology is revealing, then I will thoroughly investigate his distinction between ancient and modern technology along with his notion of enframing, after which I will then apply it to AI specifically and elaborate in what ways AI is adding to this phenomenon.

Technology As a Way of Revealing

When people are asked what technology is, usually one or two answers are given. It is either a means to an end, and/or a notably human activity. The use of hammers to build bridges, cars to quickly get from A to B, coffee machines to make coffee and smartphones to communicate with others and to search for information on the internet are all examples of human activities and these technologies are all used as a means to achieve their respective “ends”. Heidegger argues that these two definitions of technology are really the same: it is a human activity to imagine certain ends or goals and to produce and utilise the means - in this case technology - to reach them. This is an instrumental definition, and is indeed very correct according to Heidegger, even for modern technology, but still it seems to leave out something more fundamental. He states that there is a danger in not getting beyond this definition, which lies disclosed in the fact

that it allows us to think that we can fully control technology, making it seem as though the effects of technology are only a human doing. After all, aren't we the master of technology? We humans decide the ends, and simply choose when to use what technology to achieve them. (Think again of the examples given: it is us that decide to use these technologies.) This view is also named "technological somnambulism" by Pfaffenberger (1988), see table 1.1. Yet, we still seem to feel as though technology is in some sense slipping through our fingers, requiring us evermore to tighten our grip around it. Heidegger writes: "The will to mastery becomes all the more urgent the more technology threatens to slip from human control." (p. 5) According to Heidegger, we should uncover the nature of this instrumentality itself, if we are to get out of this cycle of uncontrollability: what happens when something is a means to an end? He states that instrumentality is the effecting on something, through which something else is attained. It is thus a form of causing, letting something arrive that was not present before, it is a bringing-forth: the uncovering of something that was yet concealed, it is creating. The hammer reveals a bridge that was not there before (and could not have been built without it), and similarly the car reveals shorter travel times (and thus a heightened connectivity over distance), the coffee machine reveals ready-made coffee at any time during the day and the smartphone reveals communication and the access of knowledge to be available everywhere. It is not at all always evident what it is that technological application exactly reveals. As Feenberg (2005) puts it: whereas God is a creator that truly has hierarchical relation to its creation - the world - and therefore suffers no repercussions from his creation, humans are far from being like gods in that they can only act on a system that at the same time constitutes them too. He gives an example of driving a truck carrying large quantities of metal down the highway. Though it might seem that there is an independence between humans and the world (massive loads are carried while you are listening to music safely in your truck), this independence is really only partial as you as a truck driver are stuck in traffic every day too, because of the countless other car drivers in your city. Feenberg thereby points out that technology as a means to create is not without risk of "recoil, side effects or blowback". (p. 48) Whatever is brought-forth is thus revealed, it is brought into unconcealment. Heidegger writes: "Technology is therefore no mere means. Technology is a way of revealing." (p. 11) Technology as a means reveals things that do not bring themselves forth.

The Challenging by Modern Technology

Heidegger considers ancient technology to be fundamentally different from modern technology, and since it is modern technology in particular that is concerning, we are now to zoom in on technology that has arisen from the period of modernity and onwards. The reason that he gives for this elemental distinction is the fact that the latter is based on modern science, physics most notably. Modern science and modern technology are dependent on each other: we use scientific principles to craft technological devices that were in no way possible before (e.g. the computer or the smartphone), yet our science is similarly performed by technological instruments such as microscopes, electrometers and infra-red sensors. Stating this linkage upfront makes it easy to intuit that this inter-dependence has a counter side.

Modern technology is revealing too, but it is different from ancient technology in that it is no longer merely bringing-forth, and it has taken the form of challenging instead, placing unreasonable demands onto nature. One of the examples Heidegger gives here is about agriculture. Before, the peasant would not challenge the plot of land by sowing his seeds. He simply let nature take care of the growth of his crops as he stood by. In contrary, the modern mechanised food industry desires to take the entire process of growth into its own hands, by demanding the air to yield nitrogen (a component in fertilisers), and challenging the land to produce the same crops over and over. He also argues that even a cultural symbol as large as the river Rhine can be transformed: where it can be seen more in terms of a work of art on the one hand, it is viewed only for its economic revenue value on the other. Rather than building a water or windmill along side it, part of it is dammed so that the river may reach a hydroelectric plant, reducing the river to be nothing more than a power supplier. The Rhine is demanded to supply hydraulic pressure, which can then be transformed into electric current. Similarly, coal mines are challenged to yield ore in order to stockpile it, so that it is ready whenever needed to power e.g. the steam engines in locomotives and factories. Heidegger thus argues that modern technology is more a form of challenging-forth than bringing-forth: a revealing with a demanding character. As he wrote: “Everywhere everything is ordered to stand by, to be immediately at hand, indeed to stand there just so that it may be on call for a further ordering.” (p. 17) The world is reduced to mere “standing-reserves”, which is reducing precisely because it makes it (and its parts) into a resource to be stored for later. It is no longer something that is there truly for itself, as it is now placed fully under our control. Förster (2019) describes that technology, in this regard, is thus essentially world-making. Modern technology has the aim to tame the (natural) world, because its very conception is based on the presumption that the world has to be fully predictable, observable and measurable (i.e. by modern science). A second characteristic that she adds to technology is that of invisibility. As is the case for all technology, the hammer becomes part of the body in its use - in that sense invisible - in order to be most effective: “technological processes are supposed to be opaque: They remain hidden throughout the process of usage when they function smoothly.” (p. 180)

Following this domestication of the world, even humans are now ordered to be standing-reserves. Karches (2018) elaborates that humans are made to be a resource, where they have become able to be affected by manipulations of the market economy and other external powers. This view allows for an easy comparison of human to technological capabilities, where simply the one (or thing) able to best perform its job gets to keep its position. It is by these principles that the use of technology leads to standardisation and control. Heidegger gives an example here which perfectly illustrates this: “The forester who, in the wood, measures the felled timber and to all appearances walks the same forest path in the same way as did his grandfather is today commanded by profit-making in the lumber industry, whether he knows it or not.” (Heidegger, 1977, p. 18) Humans being challenged-forth is different from the demands placed on nature, as it is we ourselves that drive technology forward. Is it then our fault that we enslave ourselves? According to Heidegger, not absolutely so. We are in a paradigm that demands us to see the world as standing-reserves. Thus, the revealing done by modern technology is no mere human doing, it is a challenging enframing that commands people to do so.

Enframing As the Essence of Technology

We have thus arrived at the essence of modern technology being enframing. As Heidegger puts it:

“Enframing means the gathering together of that setting-upon which sets upon man, i.e., challenges him forth, to reveal the real, in the mode of ordering, as standing-reserve.” (p. 20)

It is through modern science that the world is untruly pictured as a calculable whole of forces, that reveals it in a particular mode of ordering: standing-reserves. It applies a deceiving frame of reference on nature, by which we are enabled, or perhaps enforced, to perceive and act in a certain way. Rather briefly he states:

“Thus, where Enframing reigns, there is *danger* in the highest sense.” (p. 28)

This danger lies in the revealing done by modern technology: its enframing is destining or dictating. In this sense we ourselves will be taken as standing-reserve, all the while thinking that we are masters of the Earth. According to Heidegger it is exactly this ignorance that leads us to this danger: “So long as we represent technology as an instrument, we remain held fast in the will to master it.” (p. 32). This amounts to saying that we remain stuck in the paradigm of enframing such that we will be denied from the more original mode of revealing (bringing-forth). The danger of modern technology lies not in the fact that it includes lethal machines (cars, guns and perhaps also superintelligence), or that is in general much more effective than ancient technology, but rather the threat lies in the essence. Enframing resides in exactly the decision underlying this being effective “at”. Yet, Heidegger argues that this destining is not a fate beyond our control, and the question we should ask is: do we let it predominate us, or are we to become notified of this relationship we have to technology, so that we may be informed of it and in that sense no longer be directly affected by it?

AI and Enframing

As we have taken notice of the essence of technology as the broader domain in which AI is embedded, we must now identify the extent to which this conceptual landscape is applicable to AI more specifically. Immediately relevant is the fact that for AI algorithms, by and large all parameters and weights are continuously refined under an optimisation function. This can for example be seen in the Bellman equations which are used in reinforcement learning, and in deep learning where a gradient descent optimisation function stands at its very core to achieve the desired accuracy. Therefore, by presenting only quantitative output, all algorithmic returns are essentialistically constituted by the technological paradigm of efficiency in a literal way. Research by Birhane et al. (2022) on the occurrence of certain value-terms in highly influential machine learning papers reflects this too. Values such as “performance” (96%), “efficiency” (84%) are tremendously more prevalent than values related more to humans or ethics. They argue that this shows that the “technical goal” (as they call it), or being “state-of-the-art”, is taken as a

justification in itself for the beneficence of any AI research. While many of these authors claim that their mere focus on performance and efficiency is showing their algorithm's neutrality (as they are taken to be neutral values), Birhane et al. argue that these algorithms are not neutral due to the fact that they are "encoding favouritism" in that they are situated in a societal power play. The general lack of care for societal values is held to display that research is drenched by them: e.g. inequality is good in that it is supporting Big Tech's ultimate goal of performance of gaining wealth. Thus, they are arguing that values such as performance are only seemingly neutral (i.e. nonneutral) because underlying they are being influenced by biased datasets or used by centres of power. But in theory, as I have argued in the introduction, with much care and research, neutrality could be achieved. What Heidegger shows us instead is that the very fact that performance is taken to be a noteworthy value is showing the enframing at hand, and that the danger here rather lies in the fact AI could be believed to be neutral. Still they agree with the fact that the state-of-the-art label has become a method of reduction of the humane. In short, clearly enframing is indeed applicable to AI, as seen in it being an embodiment of performance and efficiency.

The Aggravation by AI

Next to the essence of AI being enframing, there are also accounts for its contribution in terms of a heightened effecting of it. Since AI is at the pinnacle of technological development, some argue that it augments this looking at the world and humans in terms of standing-reserves.

Indeed, the instrumental definition is also being applied to AI, as Förster (2019) writes. AI applications are seen as technologies that are not really intelligent themselves, because they are created by programmers to reach an end, simply by computational means. She argues that because AI systems are black boxes (although the maths may be understandable by the expert) and because its high-speed processing powers require an ever smaller physical scale, they are bringing advanced forms of invisibility: the tool becomes opaque, and the tool becomes hidden. Furthermore, she elaborates that while its capability for the prediction of human behaviour allows for an extremely smooth user experience, even the revealing of humans as mere resources stands evermore at risk.

On a different note Stolterman and Fors (2004) point out that the novel information technologies are advancing towards a fully connected world, such that individual technological applications are no longer distinguishable from each other. They argue therefore, that physical reality itself will become "intelligent", such that one is no longer interacting with just one application, but with many others at the same time, and they expect that this interconnectedness will have a colossal impact on the way that people perceive the world and experience the "good life". This especially visible in the phenomenon of smart homes. In a way this connectivity strengthens the enframing paradigm, as the environment surrounding us can be seen as fully embracing it.

Finally, due to its high level of complexity AI is able to overtake tasks that were strictly only possible to be performed by humans before, such as communication and human interaction (Vrontis et al., 2022). This comparison between humans and machines on a bigger playing field makes enframing more apparent, not only in terms of looking at the world, but also regarding

our view of humans. As Karches (2018) puts it: “We even subject ourselves to this analysis, explaining why we so readily accept the replacement of human labor by machines, despite the human consequences: if a machine can accomplish the task more efficiently, then so be it.” (p. 4)

Thus, to sum it all up, for the reason that AI is inherently opaque – a black box –, as well as the fact that AI forms a connected whole, a reality is created that sees the world in terms of standing-reserves to an increased extent, and since AI brings the prediction of human behaviour, coupled with a scene where humans are more disposable, the challenging-forth of humans themselves is intensified. These aggravations are listed in table 2.1.

-
1. Increased opacity as a tool (black boxes)
 2. Prediction of human behaviour
 3. Forming an interconnected whole
 4. Replacement of previously uniquely human tasks
-

Table 2.1

Four ways in which AI increases technological enframing.

Following this argumentation, perhaps we are not to think the danger of AI only imaginable as being maliciously against us (e.g. superintelligence), or as a tool used by some to rule over others. There is something else at hand that threatens us all. The danger is to be disclosed more in terms of a silent AI: the kind that creeps into our lives and challenges us and the world, all while we are readily accepting its use. AI magnifies the crisis of reduction by merely looking at the world in terms of resources, efficiency and productivity. It strengthens the paradigm of enframing and thus its very essence is an architecture for nonintentional dehumanisation. In the next chapter I will elaborate how this nature of AI leads to dehumanisation, and I will also conceptualise what forms it takes.

Chapter 3

The Dehumanising Character Underlying AI

The previous chapter discussed that a tendency to be reducing is immanent to AI. This reducing came in two distinguishable forms, the first being the reduction of humans to mere resources, and the second being the reduction of the way we see (and experience) the world ourselves. It was discussed that AI therefore brought a certain kind of danger through its enframing. In this chapter, I will connect this danger of reduction to the notion of nonintentional dehumanisation. After all, if we are to delineate how modern AI-applications facilitate this apparent dehumanisation, we first need a solid basis on which to lay its foundations. Therefore, it is now important to gain a clear notion of nonintentional dehumanisation. As I have discussed in the introduction, it should be strictly separated from any forms of dehumanisation that arise through a underlying power play, as assessed by the anthropological definition of technology. Crucially important for nonintentional dehumanisation is that only arises as a side-effect of a goal that is perfectly reached. The effects are not intended, in that they are not needed to reach the goal. Rather, the problem is inherent to the goal that is set, which is represented by enframed values. In the case of socially oppressive technology, this oppression can be seen as being either intended or not. The first case naturally separates itself from this notion of nonintentional dehumanisation. Even though one may justifiably note that it is an evil goal, its goal entails the dehumanising consequences, i.e. they were intended. About the second case however, it should be noted that while the negative effects arise from bias that was not intended – for example, when this bias was already present in erroneous training data (see e.g. Fountain, 2022) –, the goal was not reached, and that therefore efforts should be taken to reduce this bias, rather than to change the goal. In short, nonintentional dehumanisation is about the setting of unwanted goals all-together.

The term “dehumanisation” is a particularly ambiguous notion, as Haslam (2006) states: “the concept of dehumanization has rarely received systematic theoretical treatment. In social psychology, it has attracted only scattered attention.” (p. 1) The etymology of the word *dehumanise* reveals to us that it is composed of the Latin prefix *de-*, which signifies that that it is the reversing or opposite of the verb’s action, and the verb *humanise*, which means “make or render human” or “make humane” (Etymonline, 2022). *Dehumanise* thus amounts to meaning “mak-

ing non-humane”, i.e. the reduction of the humane character of something. Both types of the aforementioned reduction by technology and AI are connected to nonintentional dehumanisation, where the first sense is linked with a denial of humanness and the second is associated with a deprivation of humanness. This chapter will characterise how both these two forms take shape.

Haslam’s Mechanical Dehumanisation

According to Haslam (2006), most of the traditional accounts of dehumanisation do not reach beyond an intentional definition (such as the denial of another through violence and conflict), but there is also a form that is beyond cruelty (residing in the so-called “social-cognitive dimensions”) which should be incorporate too. Central to Haslam’s claim stands a denial of humanness. Now, this “humanness” is perhaps not as directly intelligible as one would like, but importantly it connects the notion of dehumanisation to a loss of what it means to be human. To account for both the intentional and the nonintentional forms of dehumanisation, he devises this humanness into two forms: 1) uniquely human characteristics and 2) human nature. The first describes the distinction between humans and the broader category of animals. It characterises them to be “cognitively sophisticated”, cultured, refined, socialised and as capable of moral sensibility. Haslam states that these characteristics seem to vary between cultures and that they may be acquired, rather than inborn, such that they are not essential, and thereby comparative. The second form is described to concern any description of the core of typical human behaviour in general. It regards the fundamental - i.e. inborn, biological, deeply rooted, unchanging - characteristics of what it means to be a human, laying in the way we are linked to the natural world. Contrary to that of uniquely human characteristics, Haslam elaborates that the human nature side of humanness is normative as it describes a typicality.

Connecting these two aspects of humanness back to dehumanisation, Haslam again composes two kinds of dehumanisation, each corresponding to a denial of one of the forms of this humanness. A denial of the uniquely human characteristics is coined by the term “animalistic dehumanisation”, through which people (especially in the case of another race or ethnicity) are considered to be like animals, which is linked with a lack of refinement, civility, cognition, morality, etc. According to Haslam, this can be linked to genocidal conflicts such as the Holocaust, but is also applicable in the context of feminism and in the disregard of the mentally handicapped. In short, it is a legitimisation for one group or person to be cruel to another. A denial of the human nature on the other hand, profiled as “mechanistic dehumanisation”, is at hand when people are disregarded in their warmth, agency, depth and emotionality, encompassing the likening of humans to machines, where only it’s superficial attributes are regarded as important. Haslam states that it is an objectification by which another’s behaviour is explained only in nonintentional causal terms.

In summary, Haslam writes: “The animalistic form of dehumanization rests on a direct contrast between humans and animals, but in the mechanistic form, . . . humans can be contrasted with machines. The shared, typical, or core properties of humanness are also those that distinguish us from automata.” (p. 258) Haslam adopts a trichotomy (consisting of humans, animals and machines) in order to explain his two forms dehumanisation by denial.

Dehumanisation As Beyond Denial

The mechanistic form of dehumanisation just described is immediately applicable to AI, as it is shaped by the comparison between humans and machine after all. Furthermore, Haslam's volitional demarcation played on the multiple accounts of dehumanisation is curiously relevant to the nonintentional entry point that this thesis takes. Mechanical dehumanisation is explained to be characterised by psychological distance, an objectification of others, or abstraction in short, rather than being moved by disgust as would be the case in animalistic dehumanisation. He is therefore right to state that "[this] abstract construal may be a cognitive basis" (p. 262), as Heidegger's notion enframing can be used to explain this. The abstracting of humans into mere resources, disregarding them as important for themselves, is exactly dehumanisation performed through a cognitive basis. The understanding of people in terms of objects leads to the denial of their subjective world, thus allowing them to be used instrumentally, as means for other ends.

Curiously, Haslam calls it objectifying, which is the reducing of any subjectivity or personal experience, but Heidegger notes that technology reduces even objects, since they too are made to become mere abstract resources. The tree is no longer a tree for itself, as it is made to be a production value representing how much wood it yields. I do not think it is necessary to dwell on the potential difference between the kinds of philosophical reduction being meant here exactly. What matters is that any form of abstraction of humans amounts primarily to mechanical dehumanisation, as according to Haslam, technology reduces humans to machines by the "robotic pursuit of efficiency and regularity, automation-like rigidity and conformity, and an approach to life that is unemotional, apathetic, and lacking in spontaneity." (pp. 253-254) Both AI research (comparing the brain to AI) and the causal explainability (i.e. predictability) of humans seem to fall under this hood.

As such, Haslam's mechanical dehumanisation provides a perfect definition for all accounts of nonintentional dehumanisation occurring through the denial of a certain humanness. It involves the way we look at humans, a denial which, to put it quite plainly, seems to have negative consequences on those who are subjected to it. Thus, of ultimate concern here is the experience of the denied, whereby the denial of humanness is a method through which ultimately a deprivation is at hand. What's more, Haslam's model of dehumanisation cannot account for forms of nonintentional dehumanisation that are not mediated by some sort of objective denial. In some cases such deprivation may be directly at hand, when for example one's experience of life itself is reduced. If we take a look again at the etymology as discussed at the beginning of this chapter, it becomes apparent that "making non-humane" does not require us to take the angle of "not granting" per se. We also need to investigate the other side of dehumanisation, one that investigates the way that we (the dehumanised) look onto life, or the world in general. As such, one is a way of looking "at" us; where we are dehumanised by denial, in that we are reduced to a machine, and the other is a way of looking "from" us; where we are dehumanised because our perspective on life is reduced. One is due to an experience that is imposed on another, and the other is about the experience that is our own. I therefore coin a secondary form of nonintentional dehumanisation: the "deprivation of humanness", which can both be mediated by the denial of humanness as well as occur without

it. The notion of denial of humanness is still relevant to identify nonintentional dehumanisation, but we must also learn to look at the latter directly. The meaning of this nonintentionality behind this deprivation of humanness and the role that technology and AI have to play in this will be addressed through Borgmann's notion of the "device paradigm".

Borgmann's Device Paradigm

Borgmann, a critic and evaluator of Heidegger's philosophy of technology examines something analogous to this secondary issue of deprivation in *Technology and the Character of Contemporary Life* (Borgmann, 1987). He too takes the perspective of how we see the world, rather how we ourselves are perceived. Technology, for Borgmann, is first and foremost characterised by availability. This availability, used to relieve us from pain and enhance our lives, should be understood similarly to Heidegger's standing-reserves. They are commodities which are ordered to stand by for whenever we find ourselves in hardship, and technology is what accomplishes this availability for us. An example Borgmann gives here is that of the availability of warmth: in modern times it is readily available by the turn of a knob, as we have a central heating system installed in our homes, whereas before warmth was something that was achieved with great effort, requiring the chopping of trees, the splitting of logs and the taking care of the fire, etc. This readiness is of course heightened by the introduction of the programmable thermostat that automatically regulates the temperature. In short, technology is relieving us from our troubles.

Borgmann goes on to distinguish "things" from "devices" in order to characterise their different relation to this availability. They are contrasted using the notion of function. A thing, as he states, is confined to its context, inseparable from the world, has a bodily relation to us and it therefore provides more than a single "commodity" or function. Turning back to the fireplace example, he writes: "It was a focus, a hearth, a place that gathered the work and leisure of a family and gave the house a center." (p. 57) Borgmann associates this manifold of function and experience with the shaping of skill, social engagement and building of character. The lack of such a nature is then corresponded with the limiting of a person's engagement with the world in full. A device on the other hand, can now be thought of as that which replaces a thing and in effect unburdens us of all the related happenings, as he argues. This unburdening is done by delivering only the fulfilment of a single function, thus also providing only a single commodity. As such, complex, high dimensional values are reduced to only singular values. Borgmann elaborates that the question of "What is the device for?" (p. 58) stands central here, in that it demands only one (relatively) static answer. Thus, he points out that their difference is that things provide multiple functions, whereas devices provide only one.

Consequently, whenever a task is reduced to a sole function, only that function is to be filled, such that a situation is created where it is the end (the fulfilment of the function) that validates the means. Borgmann's device is thus distinguished by a high plurality of means, confined only by its relatively stable function. Whatever is able to do so best gets the part. What's more, according to Borgmann the means become hidden, as it is only the end, i.e. availability, that is of importance. It is here that we can find a connection to the previous chapter. Elaborating

that modern technologies are characterised by a concealment corresponds exactly with the point advocated by Heidegger. Concerning this relationship between hidden means and availability Borgmann writes: “The concealment of the machinery and the disburdening character of the device go hand in hand. . . . A commodity is truly available when it can be enjoyed as a mere end, unencumbered by means.” (p. 60) With regards to AI, he remarks that it is really no friendliness at all, precisely because its inner workings are concealed: it is an epitome of having the means hidden to the user (see also Förster, 2019, as discussed in chapter 2). For Borgmann, machines are thus characterised by the effecting of heightened availability of the ends, through the hiding of the means.

If we are to listen to Borgmann, this replacement by reason of heightened availability goes hand in hand with disastrous consequences. About the machine he states that it “makes no demands on our skill, strength, or attention, and it is less demanding the less it makes its presence felt.” (p. 58) According to him, it is only important for the user that its functionality is familiar, the means do not matter to them, giving the example that analogue watches are replaced by electrical ones without much hesitation. Borgmann thus aims at a reduction or replacing of parts of the “good life”, and therefore he introduces the notion of the focal practice (focus means “hearth” or “fireplace” in Latin). The fireplace was thus a focal thing, as it centralised the many tasks of a household, as he elaborates. Other examples he gives are “music, gardening, the culture of the table, or running” (p. 246) and so too did the events surrounding the myths of the past categorise as focal practices; they defined what our life-world meant to us and they shouldn’t be seen as merely aiding in survival. Precisely because focal things require effort to achieve something they are readily being replaced by devices that are able to fulfil a demand more easily: this is the device paradigm.

As a result, Borgmann argues that human capabilities are overthrown by mechanical powers, social isolation has become evermore present, our relationship with nature has shifted from mastery to dominance, and even the act of working itself has become unfulfilling. Furthermore, this availability leads to discardability, as maintenance and repair become impossible or undesirable for the user, such that empty ball pens and broken smartphone components are simply to be junked and discarded. This is why he continuously refers to technology as being “debilitating”. Although the usage of a device may be justified in singular cases, depending on its context, Borgmann argues that when taking this pattern over a long period of time, it is depriving us of something nonetheless. Regarding labour he writes that although it has become easier and its output more certain, humans are left dependent, deskilled and dissociated from social and bodily engagement. The problem then, according to him, is the following: “as long as we ignore the definite ways in which we, acting technologically, have worked out the promise of technology and remain vaguely enthralled by that promise, so long simple things and practices will seem burdensome, confining, and drab” (p. 249). Taking this device paradigm by Borgmann into full account, there is something to be said for a view of technology that accentuates its tendency to be depriving us of what it means to be human: technology brings shallowness.

A Two-Fold Model of Nonintentional Dehumanisation

Connecting back to Heidegger, these two forms of dehumanisation thus arise from the results of the enframing that AI enables us to do. By viewing humans as merely being resources, we reduce our view of them, such that we deny the humanness in them. By viewing the world around us as merely being resources (seeing tasks or objects as fulfilling only single functions, rather than the complex value landscapes that surround focal things and practices), we reduce our view of life entirely, such that we are deprived of the humanness in ourselves. Or in more technical terms, the first is a looking at the human subject in itself (more objective), and the second is a looking at the human subject for itself (more subjective). Figure 3.1 displays how this conceptual arising of this two-fold nonintentional dehumanisation from AI can be visualised. To summarise, nonintentional dehumanisation takes two forms: one is caused by the making of human resources (denial of humanness), whereas the other is caused by the hollowing out of human tasks (the deprivation of humanness).

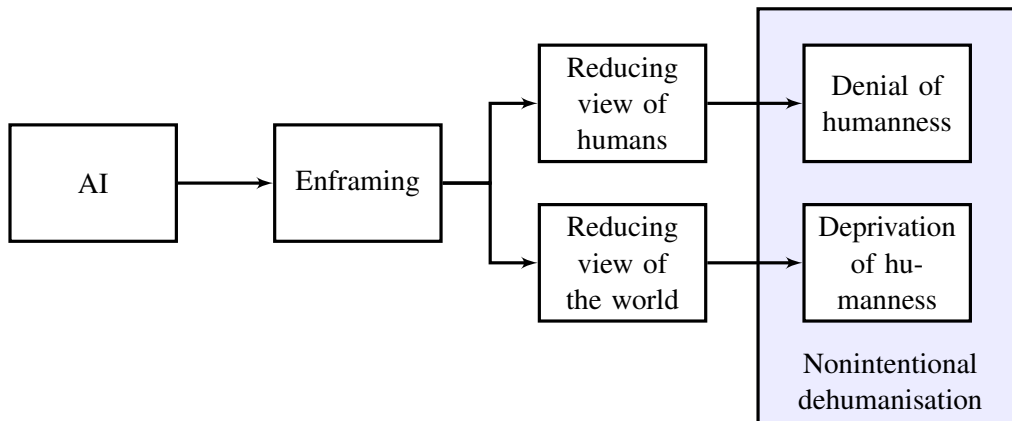


Figure 3.1

This conceptual model displays how AI leads to inherent nonintentional dehumanisation. Heidegger argued that modern technology is intrinsically enframing: a sort of lens that renders everything as mere resources in a paradigm of efficiency. As AI is embedded in modern technology, it can be argued that AI is contributing to this enframing. This enframing was furthermore argued to be reducing in two manners, on two sides of a perspective: it presents a reducing view of humans (objective side) and it presents a reducing view of the world (subjective side). Placing a reducing view on humans denies them, those who are looked at, of their humanness, whereas placing a reducing view on the world deprives those that are looking of their humanness. Nonintentional dehumanisation (the blue box) takes shape as both these forms, which will be unpacked further down below.

Borgmann's example about the replacement of a fireplace by a central heating system shows us that our humanness can be reduced regardless of whether it is first denied or not, we can also be deprived of it directly. Still, it remains necessary to keep this conception of denial of humanness too, because they are found to be occurring by a very different form of causes. The sight of another individual being instrumentalized in a system, for example, enables us to judge it as being denying that person of his or her experience of humanness. It should be noted, however, that deprivation is a lot more ambiguous than denial in its nature, both because it necessarily

emerges out of denial (thus it may come in primary and secondary appearances), and because the extent of its primary display can generally be disputed more easily. Knowing whether or not the apparent dehumanisation is by means of denial or by direct deprivation provides us not only with an easy identification of nonintentional dehumanisation for their distinct causes, but this distinction also provides us with a greater insight into finding a fitting solution to it, as I discuss in the final chapter. Both types of nonintentional dehumanisation can be inhibited by restricting the occurrence of their corresponding causes, and it is therefore that I take both types of nonintentional dehumanisation to be relevant in identifying its prevalence and in countering this prevalence in the domain of AI applications. I have listed their distinct causes and effects in table 3.1.

	Denial of humanness	Deprivation of humanness
Causes	Comparison with machines Datafication Psychological distance Instrumentalisation	Availability Hidden means (black box) Replacement of things by devices Simplifying complex values
Effects	Passification / denial of agency Lack of warmth No regard for subjectivity Denial of depth / emotionality	Deskilling Social isolation Unfulfilment Dependence Dissociation from nature

Table 3.1

This table displays the general characteristics of the two forms of nonintentional dehumanisation, which can be used for their identification in concrete applications. The left column shows the causes and effects that correspond to the mechanistic dehumanisation as described by Haslam (2006). The causes of this nonintentional denial of humanness are characterised by a general objectification of humans or a disregard for their true individuality or the individual itself. The right column shows the causes and effects that are related to the device paradigm as elaborated by Borgmann (1987). The causes of this nonintentional deprivation of humanness are characterised by a readiness of availability, the hiding of the means to attain them and this replacement-mindset that this allows for.

Chapter 4

Investigating Contemporary AI Applications

Having identified the problem of nonintentional dehumanisation as arising through the very nature of AI itself, this chapter will investigate the prevalence and appearance of this problem in contemporary AI applications. Since the areas in which AI is applied are so vast and variable nowadays, I have opted for a domain-based approach where the particular AI applications are taken as through their situatedness in a particular field, sector or domain. This allows for a descaling of the scope of this ambition, both because it changes the focus from assessing the very particularity of every AI application to more broadly defined terms, i.e. the field in which they occur, and because it permits an easy selection of what fields are explored in the first place. Two fields are investigated: 1) human resource management in the professional world and 2) healthcare. The first is selected because the making of resources out of humans seems directly apparent here, and the second is selected because it displays an infiltration by AI in a field that is characterised by human warmth, i.e. care. There are of course many other fields that also request scrutiny, such as personal life, the political sphere and democracy, brain research and anthropomorphic AI, the total size of which falls beyond the scope of this project.

Throughout this chapter, the manifestation of nonintentional dehumanisation in these fields will be identified using the proposed two-fold model from the previous chapter (see figure 3.1 for the underlying outline). Furthermore, I have argued that nonintentional dehumanisation should be understood to be occurring in two forms, the denial of humanness and the deprivation of humanness, which was split because each form primarily entails distinct and specifically related causes and effects (see table 3.1). Both the causes and the effects of these two forms are used to match contemporary phenomena of AI-induced dehumanisation, in order to gain an accurate representation of concrete nonintentional dehumanisation in AI applications that are currently used.

AI in Human Resource Management

The impact of AI-systems on the professional world is a topic often discussed. After all, the alteration of someone's job has severe consequences on their life, if only for sheer amount of the time we spend working on a daily basis. Consequences can be both beneficial, such as in the assisting with dangerous tasks, and harming, with obvious contenders such as job loss and the degradation of work quality (Selenko et al., 2022). One quickly intuits that it is a field where the introduction of AI can take many forms of nonintentional dehumanisation, since terms such as "human resources" are more involved here than anywhere else. In the near-term, the introduction of AI into the work-field is expected to have the greatest influence on mid-skilled jobs ("blue-collar work") and these occurring changes will be more about modifying only a part of a job, rather than about changing it wholly, at first (Levy, 2018). It is a vast field, composed of occupations of management (employers) - delivering the opportunities for work - and workers (employees). The manner in which this entire system of job-giving and job-taking is managed is considered to be in the domain of Human Resource Management (HRM). Naturally, since these two camps of people are affected by the HRM-strategies in action, the matter will be approached for both the employers and the employees. Characteristic of HRM is that it tends to be about enforcing strategic objectives, such as efficiently managing the efficacy of the workers (Wikipedia, 2022). According to Vrontis et al. (2022), companies are, at least to a large extent, governed by an attitude of reducing costs and improving certain functionalities and performances. Therefore, AI is incorporated as a means of increasing the effectiveness and speed of work processes. They proceed by stating that although the field of HRM has been changing since the industrial revolution, one of the primary novel deliveries of AI is its ability to perform tasks that were previously only performable by humans: e.g. communication and interaction; tasks that are key to HRM. These novel capabilities by AI portray one of the ways in which AI aggravates nonintentional dehumanisation as displayed in table 2.1.

Vrontis et al. further elaborate the exact types of AI applications used in HRM: replacing jobs, demanding human-AI collaboration, changing decision-making and the teaching of skills and task procedures to new employees, among other things. With regards to the tasks of HRM specifically (training, recruiting and performance checks), they argue that the primary activities of HRM to be infiltrated by AI are that of recruiting and performance checks. In general, their meta-analysis concludes that AI algorithms can develop a crucial role in helping humans with "better" decision-making in HR processes by reducing the time needed overall for these processes, as well as providing enhanced accuracy (at least for non-experts). In order to take a look at the experience of any current employee's, Selenko et al. (2022) propose the use of a "functional-identity model". They explain that the functional role of AI implementations in the work-field can be divided into three distinct types: it can 1) complement existing tasks, 2) dispose of existing tasks by replacing them, and 3) form new tasks all-together. Importantly, Selenko et al. note that employees gain part of their identity through the relation to their work, yielding e.g. a sense of belonging, self-esteem and meaningfulness, wherefore the way they perceive their work is ultimately affected by the way they view their work-related identity. Accompanying, they argue it to be purposeful to scrutinise

the implementations of AI in the work-field by assessing the influence of the specific manner of functional deployment, as well as the process by which this implementing happens (e.g. slowly or rapidly), on the resulting identity-changes that the workers may be presented with. For this domain I identify two distinct areas in which the implementation of AI applications allows for nonintentional dehumanisation presents to itself. On the one hand there is the side of the workers, where there jobs can be either modified or replaced, and on the other hand there is the side of management, where managerial tasks are overtaken.

Modifying and Replacing Human Jobs

One of the most upfront changes that can be imagined in the field of HRM is the direct replacement of human workers by their mechanical counterparts. In this case, a job is disposed of entirely, and the individual is rendered without work, such that their identities may be challenged, as is advocated by Selenko et al. (2022). Such a direct replacement can be directly seen to be a comparison with machines, i.e. a cause of denial of humanness. They proceed by noting that such a radical replacement does not only affect those who are replaced; it is equally important to understand how such a replacement leads to resulting dynamics that stretch far beyond the influence sphere of a single individual. Such changes may alter the entire “social fabric” of the jobs of any co-workers. This may reduce the the positive part of their identities too, such that in this roundabout way, others are deprived of their humanness as they are lead towards social isolation in the workplace. It is therefore, that Selenko et al. argue that it is more desirable to opt for a general attitude of AI augmenting human task performance (preferably leading to identity-enhancements) over that of replacing human jobs entirely (as it challenges their identity, and other’s, the most). Merely complementing existing tasks entails that in a way a worker’s identity does not necessarily have to be threatened at all; it can also be enhanced by e.g. the removal of tedious or repetitive tasks from their responsibility. Since the replacement of humans is thus immediately seen as dehumanising (denial of humanness) Raisch and Krakowski (2018) state that it is very characteristic in the literature of HRM to have a sole focus on the promise of augmentation, rather than automation. This is because, as is generally conceived, automation entails the replacement of jobs, whereas augmentation has the opportunity to be beneficial the the workers too. This tendency is visible in some of the views just discussed, in that it was stated that it is both expected and desirable that only parts of jobs are changed, rather than replacing them wholly. Vrontis et al. (2022) opt for a “conflict resolution model”, for example, that draws on the values of efficiency and equity and attempts to find a balance between them. In this way they hope that we remain in control such that we we are able to enforce solely the positive effects of AI in HRM decision-making, while ruling out the factors of privacy infringement, bias and the potential for domination. So too, do (Selenko et al., 2022) argue that if we wish to protect employees against the largest identity challenges, we can opt for an easily intuited all-solving policy that restrains the full automation of tasks and therewith only supports human labour (and therewith their identities) instead.

Yet, Raisch and Krakowski (2018) warn us of the fact that these two processes of augmenta-

tion and automation are heavily intertwined; a so called “augmentation-automation paradox” is at hand. According to them, the impacts of a merely augmenting AI application cannot be fully grasped without understanding that the augmentation of a task at a lower level leads to a conception of more efficient automation at a higher level. Similarly, the opposite is equally true, where conception of task being fully automated should be seen as being augmenting to human labour in its implementation somewhere along the way in the management hierarchy. It is important to note here though, that the term “augmentation” does not necessarily mean that it is “better” for those who are affected: one should therefore re-conceptualise this paradox in more value-neutral terms, such as a “modification-replacement” paradox. Proceeding, as Selenko et al. (2022) argued similarly, Raisch and Krakowski therefore note that the effects of the automation of a job entirely has consequences that reach much farther than just for that replaced person, such that management positions that were normally social could turn out to become professions of merely interacting with machines instead. This is a deprivation of humanness, since in essence it entails the reduction of a complex value-landscape, namely: social interaction. Equally, the augmentation of a job amounts to both automation on a lower level (e.g. a single task of a job is automated) and to a conceptualisation of this AI-worker system as automation on a higher level. The former has the potential to lead to a deprivation of humanness, as it secretly entails the hiding of means (workers are distanced from their products), which might amount to deskilling and feelings of dissociation and unfulfilment. The latter on the other hand, allows for the possibility of denial of humanness, and it is here that we proceed to AI-based management itself.

Algorithmic Management

With this inter-relatedness between augmentation and automation in mind, Raisch and Krakowski take a step to address its influence in the sphere of HRM: “Managers limiting their perspective to either automation or augmentation risk developing partial and incomplete managerial solutions.” (p. 18) They hold that it is therefore naive to hold that management can remain to be run exclusively by humans, if only we restrain from the use of automation and focus solely on the use of augmenting AI. Consequently, they proceed to elaborate that AI management is different from human management in that it leads to very consistent behaviour (“formal rationality” as they put it). However, it is argued that what is most principally new is that AI management essentially only focuses on maximising output (see also chapter 2), and that it is thereby directing and constraining human behaviour. The human form of management, on the other hand, is argued by Raisch and Krakowski to be characteristic for our capability of making decisions based on a reflection of a complex multiplicity of values, as well as not being as liable to statistical bias (although other biases are of course present). This transition from human to algorithmic management, which brims with maximalisation, thus characterises a psychological distance to the workers, as well as their instrumentalisation, such that we can speak of a denial of humanness here.

Nonintentional dehumanisation can also be seen in the HRM-sphere with e.g. the rise of the recruitment algorithm. After all, one of the most central roles of HRM/employers is finding new employees. According to Fritts and Cabrera (2021), there is much worry under HRM profes-

sionals that the use of these algorithms might lead to dehumanisation of the applicants, yet survey responses showcase that it is booming business nonetheless, as over half of the respondents state that they expect that AI will be incorporated in the hiring process within a few years. As performance, accuracy and efficiency form the backbone of any company-level decision, recruitment officers opt for the use of AI in the recruitment process in order to optimise finding the applicant with the highest prediction of expected sales and general performance, as well as decreasing the time spent on searching for candidates. Fritts and Cabrera state that this is attained by means of e.g. personality tests, gameplay, puzzle solving and video interview analysis, and that HRM even has the ability to tweak these programs so that the finding of a fit that is able to fulfil the specific demands of the job is maximised with additional precision. They expect that the software will be applied to a large extent, especially in the earlier stages of the recruitment procedure (e.g. sourcing of potential candidates and screening them in terms of personality and fit with the organisational culture), often overtaking parts of the task, or replacing it wholly entirely. To top it off, even the actual interview could be done by AI-agents (or run as a parallel analysis in the background), leaving only the final selection phase up to the human recruiters, as they argue that the more high-end human judgement is as of yet unattainable by AI decision-making.

With regards of the consequences of these algorithms, Fritts and Cabrera argue that it is by looking at the applicant-employer or employee-employer relationship that its accompanying risks come to light. Although a company has obligations to its stakeholders, rather than just its shareholders (this adds the employees), wherefore a good employee-employer relationship is of paramount importance to the company, the e-recruitment is expected to change this already adversarial relationship for the worse nonetheless. Importantly, they state that “[t]he sense of dehumanisation at play in the case of hiring algorithms does not involve conceiving of other humans as subhuman, but instead involves removing the human presence from some process or domain”. (p. 794) By drawing on Nguyen’s analysis of “artificial values” they identify the crucial reason for this to be the fundamental difference between the values of a human recruiter and the values embedded in these algorithms. They argue that recruitment algorithms are therefore morally objectionable because its implementation changes an activity that was normally an intricate yet ambiguous interplay of many values to be now governed by values that are artificial. In stead of being valued as a real, particular human being by the judge you are evaluated by an algorithm that looks at you only in terms of data points, such that those subjected to it feel that the artificial judge is inhumane, and that any victories feel hollow since you have no longer truly convinced anyone with the displaying of your being. As such, Fritts and Cabrera state that although humanisation may come with bias, the ease or promise of algorithmic objectivity of bringing quick, predictive, incredibly complex, superbly accurate and objective judgements seems to come at the price of being inherently dehumanising. Thus, it is not privacy, algorithmic bias or the feeling that one is being objectified that is the biggest worry of these artificial recruiters, but rather that it gives rise to the feeling that the human touch is lost. The nonintentional dehumanisation found here is more due to deprivation than denial: the employee-employer relationship can be likened to a focal practise by Borgmann (1987), in that it is a complexly valued practise that replaced by an artificial one.

AI in Healthcare

The second area of investigation is that of Healthcare. Perhaps more so than in other professional fields it is important for the field of healthcare that there is a presence of a certain humanness. But, what differentiates healthcare even more on this topic of dehumanisation is that, paradoxically, it presents situations in which dehumanisation can even be good or even desirable, as when it is necessary to inflicting pain during an operation (Dawson, 2021). For precisely this reason, Palmer and Schwan (2022) separate functional from non-functional dehumanisation. The functional form is characterised by a distancing of oneself from an emotional view of a patient for the reason of achieving higher “medical values”, e.g. considering a body to be lacking subjective experience when one has to make an incision, or likening a body to a machine for reasons of diagnosis. In this functional sense, it is thus possible that the denial of humanness, as caused by e.g. psychological distance, comes to light in a way that is beneficial. The non-functional form is described as the decrease of care that is not supplementary to a desired “medical end”, such as the depersonalisation of care. In fact, this allowed-for form of dehumanisation is necessarily intentional, as it is used as a means to give rise to something else. Thus, functional dehumanisation can never be nonintentional. This distinction is important to keep in mind, as it allows us to identify the truly nonintentionally dehumanizing AI applications (the non-functional type) in healthcare.

Healthcare too is a domain to be heavily influenced by the introduction of AI applications. Davenport and Kalakota (2019) identify the key aspects of its utilisation to be “diagnosis and treatment recommendations, patient engagement and adherence, and administrative activities.” (p. 94) For each of these tasks AI is already on par with human performance and in some occasions even superior to it. They explain that at the moment of writing, they do not expect AI to be replacing jobs in their totality, due to the narrowness of the scope of these algorithms, but rather that they will change only parts of it. As they state, algorithms detecting malignant tumors may for example outperform radiologists in their raw capacity to screen images, but it is important to note that these radiologists have many more responsibilities, such as consulting with their colleagues in their diagnosis and treatment plans as well as having to perform the actual medical interventions, among many other activities: most of all they are the ones connecting the dots. Yet, I want to point out that this restraint is both likely to lift in the future, as it is only a matter of time before AI systems become broader, and it should be noted that despite of the narrowness of a specific AI application, its implementation can have dehumanizing effects on a system that far surpass its sole role of usage (analogous to the “augmentation-automation paradox” by Raisch and Krakowski, 2018). I identify two main areas in which nonintentional dehumanisation occurs in healthcare by means of the implementation of AI, which again entail that of management or governance on the one hand, and the replacement of doctors on the other. The former refers to e.g. governance by electronic health records and derogatory medical terminology that highlights the way that patients are viewed, whereas the latter refers to the actual replacement of human caretakers and human contact in healthcare (such as in primary care and psychiatry).

Algorithmic Healthcare Governance

Similar to the field of Human Resource Management, the manners in which healthcare is done is also propelled by reasons of effectiveness and efficiency, such that it is crucial not to see AI merely in a light of neutrality. Karches (2018) therefore argues that we should scrutinise whether AI applications in healthcare provoke closer or more removed examinations of a patient's body. To illustrate the difference, he compares two tools used in healthcare: a stethoscope and the Electronic Health Record (EHR), explaining that there is indeed a major difference between the two. The use of a stethoscope is likened to the use of a hammer (as discussed in chapter 2), where its use allows the caregiver to "inhabit" it in order to reveal the patient's body for closer examination, i.e. as it truly is. Thus the stethoscope can be seen as an improvement to healthcare. The EHR, however, is regarded as delivering no such thing, as in fact it leads to a heightened distance between the physician and the patient. He elaborates that with the EHR a drastic depersonalisation is at hand, because patients are clustered by their statistical resemblance in order to create so called cohorts that are constituted by similar combinations of age, sex, race and other commonalities (even DNA is not beyond the scope here), such that precise advice is derived from demographic data, rather than the unique individual, which is used to mandate what group (and thereby the individual) is to be screened at what time and for what disease. Consequentially, through such a record system the actual personal circumstances are left out and the individual is reduced to merely its most basic medical characteristics, i.e. a problem-set to be solved, as Karches puts it. The patient is put through a datafication process entirely (a cause of the denial of humanness), and the availability of this data allows external powers (he exemplifies the state or insurance companies) to transform healthcare to that of a paradigm of efficiency. Since efficiency is constitutional to AI, Karches remarks that the EHR will only be able to focus on demanding a societal riddance of disease, such that no listening ear is turned to the actual uniqueness of a particular patient's problems. This allows for a dynamic that values statistical care over individual care. The only thing that is considered to be important is statistical significance or "healthcare in general". As such, this datafication of the patients decomposes the actual to mere quantitative components, which in turn allows for their "curing" to be used in an instrumental fashion. Both this datafication and this instrumentalisation are forms of denying the patients of their humanness. Whats more, Karches even links the EHR to a resourcification of the physicians themselves (denying their humanness too), allowing for them to be ordered by external economic incentives so that healthcare threatens to become "care on a production line". (p. 7)

The dehumanisation by the depersonalisation of care becomes even more apparent when looking at the medical terminology that this governance facilitates. According to Dawson (2021), the language used to refer to patients imposes the risk of not seeing them for the humanity that they present. Derogatory terms such as "bed blockers" (staying your appointed care-time overdue) and "medically optimised" (no longer requiring attention as you are found to be too healthy for care) are therefore denying patients of their humanness in that they are disregarded for their inner life. The example that Dawson gives is that the term "medically optimised" tends towards a sort of emotional distance in that it assumes a passivity of the patient as they are viewed as

“non-persons”, viewing them merely as being objects in a system, and states that it would be better to refer to them as being “ready to consider discharge” so that they remain their capacity for their own agency. Considering the role that AI has to play in the gathering and analysing of the patient’s data that makes these terms emerge, it is easy to see how in this sense, datafication is inherently tied with psychological distance, rendering a denial of humanness in close proximity. Dawson mentions that deindividuation is further apparent because such phrases reduce people to an illness, glossing over their particularity and shrinking them to be “part of a herd”.

As such, AI tools such as the EHR can lead to nonintentional dehumanisation in the entire clinical attitude, both expressed and fostered by such depersonalising terminology. The insightful remark by Dawson that the system is prioritised above the individuals shows us exactly what the term “governance by AI” means: it is a shift in the overarching managerial perspective that as it were takes the "managing" as more important than the "managed". As was also the case with HRM, governance signifies a denial of the humanness of those governed: mainly the patients, but also the caregivers.

Replacement of Human Caretakers

With the introduction of narrow AI systems to healthcare, one of the most human tasks of care is that which include communication with patients. As I have discussed in chapter 2, one of the primary novel dangers of AI is its ability to replicate this human communication. As I have just discussed, the utilisation of AI for documentation (such as the EHR system) stands at risk of presenting a governance by efficiency and thereby enabling dehumanisation. But, with the replacement of actual human caregivers (and thereby human-empathetic care), consequences might be arising that are of a different order entirely. Light should therefore be shed the possibility of this replacement in the domain of healthcare, however unlikely it may seem that AI is able to convince us of its qualities.

Such communicative forms of human empathetic care are visible in for example primary care and nursing homes, but perhaps it is most apparent in the field of psychiatry. In a survey conducted by Doraiswamy et al. (2020), it became clear that almost half of all psychiatrist respondents believed that AI would have little to no influence on psychiatry in the next 25 years, and only about 3.8% of the respondents believed that a total elimination of human psychiatrists was apparent in the future. In general, it was expected by most that tasks such as documentation are very likely to be outsourced to AI, but the ability of AI to perform empathetic care was perceived to be very unlikely. Doraiswamy et al. state that as of today, self-care and self-monitoring by means of AI is evermore present. Equally, it is used for “psychiatric diagnoses, symptom tracking, disease course prediction, and psychoeducation” (Pham et al., 2022, p. 1). There are indeed some explorations covering the advantages of the replacement of human caregivers by AI. Pham et al. explains benefits such the use of chatbots in psychiatry to teach emotional coping mechanisms, and to provide help to those struggling with communication problems (e.g. trouble recognising facial queues). These could for example be attained by chatbots (which are argued to be more engaging than e.g. a book) and phone trackers (which could be used to identify manic episodes from

e.g. typing-behaviour at all times). Ultimately central stands the advantage of being absolutely unjudgemental. On a similar note, Palmer and Schwan (2022) considers the beneficence of using AI in cases where “shame-induced barriers” gives rise to a hindrance of stepping up to a real human for care, or whenever human care provided – cases such as feeding, going to the bathroom, changing sheets – is inducing feelings of shame. Here too AI is argued to be a valid alternative for human care or perhaps even an upgrade.

As such we have established that it seems possible that there are ways in which even the human attentiveness of caring could be regarded as redundant. Nevertheless, the manner in which these algorithms are actually helping those who need care remains unclear. Karches (2018) writes against the claim that we should replace physicians because AI algorithms – or “iDoctors” – as it is found that they “routinely outperform practitioners’ clinical intuition” (p. 2), along being more cost-efficient. Such measure of surpassing of human capabilities is taken to be too simplified what healing or caring really is, as he argues healing to be a highly ambiguous practise in reality. The comparison between doctors and machines is happening on the basis of certain pre-selected medical values, about which he states that “the iDoctor’s reduction of human beings to a set of medical concepts necessarily fails to understand life as it truly is, in its fullness.” (p. 8) According to Karches, there are often trade-offs to be made between multiple “goods”, and the many accounts of symptoms and side-effects by the patient need to be taken into consideration. Karches links primary care to Borgmann’s focal practise and explains that this complex value-landscape, or ambiguity, displays that it is not for the riddance of disease that people go to physicians, but that it is primarily for a relief of suffering that people seek healthcare. This is a point that is also elaborated by Jotterand and Bosco (2020). Rather than a maximalisation of specific medical values, empathy and true understanding seem to be whats most important in healthcare: something which AI cannot attain, given its nature. AI care is only able to support enframed care-values, but whenever empathy is needed it remains absent. Palmer and Schwan admit that the use of AI to avoid feelings of shame may pose the risk of legitimising our shame mindset about these healthcare problems, and as such we might state that vulnerability is important if one desires to move beyond having shame for his or her condition.

In short, these quantitative values are reached, yet the sole attainment of this goal elicits non-intentional dehumanisation. Thus, the replacement of the human touch in healthcare is liable to lead to a simplification of complex values, with reveals to us once more that there is a deprivation of humanness at hand. It is always a trade-off between raw recovering (e.g. a broken body that needs to be “fixed”) and the improvement of well-being, but in all areas where well-being is considered to be important this principle of AI needs to be taken into account. The very fact that we believe AI to be able to provide care shows that even we ourselves are stuck in the paradigm of efficiency and therewith exposing patients to nonintentional dehumanisation.

Chapter 5

Discussion

Overview

In the previous chapters I have first followed Heidegger's definition of modern technology, which holds that technology is not neutral or a simple means to an end, but rather that it is enframing: a paradigm that reveals things as mere standing-reserves, i.e. resources. AI takes part in this enframing in that the output of its algorithms are literally constituted by a quantitative maximalisation function. Furthermore, AI was argued to aggravate the problem because it brought increased opacity (it is a black box), the prediction of human behaviour, the formation of an interconnected whole that embraces this enframing and because it is able to replace tasks that were previously uniquely human (communication and interaction). The underlying danger of this nature of AI was elaborated to be nonintentional dehumanisation: dehumanising side-effects that arise only upon the perfect attainment of a predefined goal. This dehumanisation comes in two forms: the denial and deprivation of a certain humanness. The denial of humanness, caused by the comparison of humans to machines and datafication, induces passification and no regard for the subject's feelings. The deprivation of humanness on the other hand, caused by the over-availability, the hiding of inner workings of machines and the simplification of complex values, gives rise to the deskilling, and feelings of dissociation and unfulfilment in humans. Thus, the readily acceptance of AI due to its effective goal-reaching capabilities is connected with danger, as it has the potential to bring about a disastrous aftermath.

Finally, this theory was formed into a model (see figure 3.1), covering both the forms of denial and deprivation of humanness in terms of their respective causes and effects (see table 3.1), and using this model the corresponding negative consequences were examined in two domains of AI applications, HRM and healthcare. The resulting causes of each of the distinct types of nonintentional dehumanisation are organised by the type of change that AI applications bring in table 5.1. It can be seen that it is mostly by means of a direct replacement or by means of management/governance by AI that a denial of humanness is brought forth. The deprivation of humanness seems more directly connected with the modification of jobs and the interaction with machines (e.g. when one's doctor is a robot). The deprivation in a certain people that arises as a secondary form to a denial of them first is not explicitly mentioned, but it should be noted that

in any social context, the denial of the humanness of certain individuals (e.g. by replacing them) leads to the remaining people to be deprived of their humanness, as they no longer have as much human connection. This is especially visible on the workplace. In total it might be said that the more governance related AI changes lead to a denial of humanness of those who are subjected to it, and that the removal of a certain human touch leads to a deprivation of humanness of those who interact with these AI applications.

Domain	Change by AI	Denial of humanness	Deprivation of humanness
HRM	Replacing jobs Modifying jobs	Comparison with machines	Interacting with machines only Deskilling Dissociation and unfulfilment
	Algorithmic management	Maximising worker output	Hollow victories Adversarial relationships
Healthcare	Algorithmic governance	Statistical over individual care Denial of individual agency Instrumentalisation of doctors	
	Replacing doctors		Lack of empathetic care No true vulnerability

Figure 5.1

Summary of the distinct forms of nonintentional dehumanisation that arise through AI-induced changes in the domains of HRM and healthcare. For every type of change that AI brings to these domains, the causes that relate to the distinct forms of this dehumanisation (the denial and the deprivation of humanness) were identified. The results are listed in the two right-most columns respectively.

Limitations

It is only natural that this thesis has certain limitations. First and foremost I will address the the research done to identify nonintentional dehumanisation in the contemporary domains of AI applications in chapter 4. While the main premise of this thesis requests a scrutiny of the effecting nature of the AI applications in all domains, the finite nature of this project enforced the decision to investigate only two contemporary fields. It can easily be imagined how nonintentional dehumanisation is also likely present many other domains, such as in personal life (e.g. the emergent smart homes and their influence on our agency and independence), the potential replacement of education (deskilling), the effects in the political landscape, but also in the field of neuroscience, where the expansion of our understanding of ourselves is often used as an argument to conduct AI research. What is the influence of the creation of such human-like AI agents on the conceptions that we have of our cognition and behaviour? Of course the two domains that were investigated are also changing rapidly as we speak, and thus changes in the exact consequences are likely to alter or become clearer through time. Moreover, the model of identification itself (i.e. the causes and effects from table 3.1) can likely also be improved with more deliberation and analysis. Thus, if we are to grasp the full phenomenon, more research needs to be conducted in order to fully explore all the causes of nonintentional dehumanisation and furthermore its occurrence should be

inquired in all of the contemporary domains of AI implementations.

Secondly, one can also have theoretical objections concerning the central propositions of this thesis. For example, the exact meaning of “humanness” can be contested: what does a good life entail, and can we agree on its definition? Furthermore, one can also oppose that AI does not have to be concluded as being dehumanising in its totality, when taking into consideration that it may also bring novel forms of humanisation instead. In fact, even Borgmann (1987) elaborated that while the emergence of new focal practices through technology was unlikely, technology still provides us with a basis that can contribute to existing focal practices by delivering the time, instruments and equipment to be able to attain them, such as one may argue for e.g. modern camping gear. He states that whereas Heidegger points at the pretechnological world to arrive at focal things, technology needs to be recognised for this possibility to allow for greater contact and engagement with the world. This positive side of humanisation of AI applications needs to be weighed against its potential side-effects of nonintentional dehumanisation and likely the answer will lie in an intricate balance between the presence of the former and the absence of the latter. Perhaps we should see some of the inherent dehumanisation as an evil that is necessary for achieving greater ends, such as is argued about the functional forms of dehumanisation as necessary in healthcare (Palmer and Schwan, 2022). In short, if we are to circumscribe the exact “problem” of nonintentional dehumanisation more accurately, we should reflect on this critique in order to refine and meliorate the proposed conceptual model.

Future Implications

With regards to the future, there are two fundamental questions to ask if we aspire to overcome or put a hold to this problem of nonintentional dehumanisation in AI applications. The first to asks is: “Can we do something?”, and thus we need to ascertain whether we are in control over AI. The second question to ask is: “What can we do?”, such that we are in need of ways to move forward. These two questions are addressed now.

On the Ambiguity of Human Control

Pivotal to understanding our responsibility of the effects of technology, and to the creation of guidelines for the future, stands the ambiguity of our control. In chapter 1 I have elaborated four different views on our usage of technology (See table 1.1), and as this thesis departed from the conception of technology as enframing, as put forward by Heidegger (1977), I correspond my view of our control to his. After all it is we ourselves that take part in the development of new technologies and AI applications. Without us, it has no incentive to advance. Heidegger also notes, however, that we are not fully in control either, so long as we do not have a free relation to it. As I have explained in chapter 2, he regards technology as a lens that dictates us to view the world and ourselves in such a way as to enable a challenging of both. In that way, technology is a lens through which we tend to look in a reducing way, which can theoretically be undone. One should therefore not speak of a deterministic relation, but rather about one of ignorance.

The anthropologists indeed share in the quest to contest this ignorance, but the only difference is that the model from this thesis aims to present an alternative reason for much of the emergent dehumanisation that is of a nonintentional type. Perhaps we can even say that some of the naivety found in the short article by Davis (1974) in the introduction is not so wrong after all, as it is us humans that need to reflect on our implementation of technology. In conclusion, if we are to delineate ways to combat this nonintentional dehumanisation in AI usage, we should certainly look at the possibilities residing in the domain of our control, and we should be forever weary that the constitutional nature of AI itself may lead us astray.

The Way Forward

Looking ahead, the primary message that this thesis presents for the future is that if we are to combat this paradigm of nonintentional dehumanisation, we need to be wary of the values that are inherent to the AI-induced changes that we tend to readily accept. Importantly, the line between good and bad AI applications is thin, and the repercussions of the usage of each specific utilisation needs to be carefully identified. AI applications often have both positive and negative sides to them, such that the intricate balance between them should be sought after. Unfortunately, backlash has the nature of only occurring afterwards, rendering timely recognition difficult. Therefore, the proposed model that differentiates the causes and effects (see table 3.1) could be used as guidelines to identify a possible onset of danger in new applications. As I have discussed in the limitations section, Heidegger would point us back towards the usage of technology that is merely bringing-forth, rather than challenging-forth, and thus aims at general retreat to premodern technology. But, we must ask ourselves, can we and do we want to go back that far? Do we wish to give up the immense technological and scientific advancements in crucial sectors such as healthcare and agriculture, and furthermore, do we wish to abstain from the general ease that technology brings to accomplish all kinds of tasks easily for us? Of course one cannot forget that with the rise of modern technology, new forms of buttressing the very humanness that we wish to protect were made possible. AI will likely only enhance these fortunate properties and thus it would be more favourable to demarcate the good from bad applications by the means that we do have, rather than to abstain from the use of AI entirely.

As I have already stated, the knowledge of which of the two forms of nonintentional dehumanisation is at hand enables the potential for an impediment of its corresponding causes in new applications. With regards to the causes of denial of humanness, Haslam (2006) argues that empathy is the solution out of his mechanistic dehumanisation, as he states that the lack thereof is facilitating an objectified perception of an individual. By reducing one's indifference towards others and thereby re-individuating them, the violation of many of the aforementioned causes of denial is easily avoided. Fritts and Cabrera (2021) opt for a Kantian disavowal of the instrumentalisation of the applicants that are subjected to recruitment algorithms, which basically amounts to a similar view of one another, in that it is important not to use someone as mere means, and to have respect for the persons as such instead. With regards to the deprivation of humanness, I have stated that it is much more ambiguous than its denial counterpart, because in several ways it is

tied to this denial, and the truth of this deprivation is also more liable to be disputed, as it depends on one's perspective. In fact, since technology also has the opportunity to improve our humanness, i.e. provide the good life – thus elevating our world-view and (re-)humanising us –, the causes of this deprivation form of nonintentional dehumanisation are a lot more ambiguous as to whether they are strict causes at all. Exemplifying a possible solution here, Jotterand and Bosco (2020), explore ethical guidelines for healthcare applications, such as the promotion of human interaction (an axiom of the good life that stands firmly embedded in e.g. Maslow's hierarchy of needs). They argue that if we are to make sure that human users remain in control of their lives, AI algorithms and their effects should be kept utterly transparent to the user. In total, the causes of denial seem to be easier to address than those of deprivation, as the way that AI allows us to treat human beings is more intuitive to counter than the way that AI as a whole may be degrading our perception of the world.

Conclusion

This thesis has argued how the use of AI gives intrinsic rise to nonintentional dehumanisation, which was connected to Heidegger's essence of technology: enframing, a paradigm of efficiency that enables a viewing of the world and ourselves merely in terms of resources. This nonintentional dehumanisation was specified as those dehumanising side-effects that consequent when an AI application perfectly reaches its goal. Hereby this nonintentional dehumanisation is starkly different from the dehumanisation that is put forward by a more anthropological view of technology. Furthermore, it was elaborated to take shape in two forms: the denial and the deprivation of a humanness. As such, a model was created that contained the respective causes and effects of each of these sub types, such that we could therewith identify the occurrence of this nonintentional dehumanisation in both existing and new AI applications. Investigation of two contemporary domains – that of HRM and healthcare – showed that AI change that is governance-based was associated more to a denial of humanness, whereas change that involved a dismissal of a human touch was related more to a deprivation of humanness. Finally, it was argued that we are ultimately in control of AI usage and if we wish to combat this paradigm of nonintentional dehumanisation, we need to sharpen our judgement concerning our readily acceptance of AI solutions.

Acknowledgments

Thanks to dr. Olivia Guest for her excellent supervision and support with e.g. literature handholds during the writing of this thesis, and to my thesis group for their fine counsel during the weekly sessions.

References

- Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., and Bao, M. (2022). The Values Encoded in Machine Learning Research. *ACM International Conference Proceeding Series*, pages 173–184.
- Borgmann, A. (1987). *Technology and the Character of Contemporary Life*. University of Chicago Press.
- Davenport, T. and Kalakota, R. (2019). The Potential for Artificial Intelligence in Healthcare. *SSRN Electronic Journal*, 6(2):94–98.
- Davis, R. (1974). Technology as a deterrent to dehumanization. *Science*, 185(4153):737.
- Dawson, J. (2021). Life & times medically optimised: healthcare language and dehumanisation. *British Journal of General Practice*, 71(706):224.
- Doraiswamy, P. M., Blease, C., and Bodner, K. (2020). Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artificial Intelligence in Medicine*, 102(July 2019):1–7.
- Etymonline (2022). dehumanize (v.).
- Feenberg, A. (2005). Critical theory of technology: An Overview. *Tailoring Biotechnologies*, 1(1):47–64.
- Förster, Y. (2019). Artificial Intelligence: Invisible Agencies in the Folds of Technological Cultures. In Sudmann, A., editor, *The Democratization of Artificial Intelligence*, chapter 11, pages 175–188. transcript, the democr edition.
- Fountain, J. E. (2022). The moon, the ghetto and artificial intelligence: Reducing systemic racism in computational algorithms. *Government Information Quarterly*, 39(2):101645.
- Fritts, M. and Cabrera, F. (2021). AI recruitment algorithms and the dehumanization problem. *Ethics and Information Technology*, 23(4):791–801.
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3):252–264.

- Heidegger, M. (1977). The Question Concerning Technology. In Vattimo, G. and Valgenti, R. T., editors, *The Question Concerning Technology and Other Essays*, chapter 1, pages 3–35. Harper Torchbooks.
- Jotterand, F. and Bosco, C. (2020). Keeping the “Human in the Loop” in the Age of Artificial Intelligence: Accompanying Commentary for “Correcting the Brain?” by Rainey and Erden. *Science and Engineering Ethics*, 26(5):2455–2460.
- Karches, K. E. (2018). Against the iDoctor: why artificial intelligence should not replace physician judgment. *Theoretical Medicine and Bioethics*, 39(2):91–110.
- Kerr, A. D. (2021). Artificial Intelligence, Gender, and Oppression. In Filho, W. L., Azul, A. M., Brandli, L., Salvia, A. L., and Wall, T., editors, *Gender Equality*, pages 54–64. Springer.
- Khole, C. and Hande, A. (2018). Artificial Evil Intelligence and Ethical Approaches in Development. *International Research Journal of Engineering and Technology (IRJET)*, 5(10):1581–1584.
- Leufer, D. (2020). Why We Need to Bust Some Myths about AI. *Patterns*, 1(7):100124.
- Levy, F. (2018). Computers and populism: Artificial intelligence, jobs, and politics in the near term. *Oxford Review of Economic Policy*, 34(3):393–417.
- Mosco, V. (2004). *Myth, Power, and Cyberspace*. MIT Press.
- Munn, L. (2022). *Automation is a Myth*. Stanford University Press.
- Murphy, J. (2018). Artificial Intelligence, Rationality, and the World Wide Web. *IEEE Intelligent Systems*, 33(1):98–103.
- Palmer, A. and Schwan, D. (2022). Beneficent dehumanization: Employing artificial intelligence and carebots to mitigate shame-induced barriers to medical care. *Bioethics*, 36(2):187–193.
- Pfaffenberger, B. (1988). Fetishised objects and humanised nature: towards an anthropology of technology. *Man*, 23(2):236–252.
- Pham, K. T., Nabizadeh, A., and Selek, S. (2022). Artificial Intelligence and Chatbots in Psychiatry. *Psychiatric Quarterly*, 93(1):249–253.
- Raisch, S. and Krakowski, S. (2018). Artificial Intelligence and Management : the Automation-Augmentation Paradox. *Academy of Management Review*, pages 1–48.
- Segal, R. A. (1998). *Encountering Jung on Mythology*. Princeton University Press.
- Selenko, E., Bankins, S., Shoss, M., Warburton, J., and Restubog, S. L. D. (2022). Artificial Intelligence and the Future of Work: A Functional-Identity Perspective. *Current Directions in Psychological Science*, 31(3):272–279.

- Stolterman, E. and Fors, A. K. (2004). Information Technology and the Good Life, Information Systems Research: Relevant Theory and Informed Practice. *Information Systems Research. IFIP International Federation for Information Processing*, pages 687–692.
- Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., and Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *International Journal of Human Resource Management*, 33(6):1237–1266.
- Wikipedia (2022). Human resource management.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile Books.