

**Reframing the Moral Status Question: An Investigation into the Moral Patiency of
Intelligent Technologies**

Cassandra Beyer

Student ID: 8424650

Frankfurt School of Finance and Management

Bachelor Thesis

Bachelor of Science (B.Sc.) in Business Administration

First Supervisor: Prof. Dr. Sebastian Köhler

Second Supervisor: Prof. Dr. Peter König

15 Jan 2024

Text of Certification

I hereby confirm that the Bachelor Thesis presented by me has been prepared independently, using no other sources, resources and other aids than those mentioned. All parts – literally or by their meaning – taken from published or non-published sources are credited as such. The Bachelor Thesis in its current or similar form has never been submitted as a graded assignment.

Date: 15.01.23

Signature

Table of Contents

INTRODUCTION	4
SCOPE AND OBJECTIVE.....	7
THE MORAL STATUS DEBATE THUS FAR	9
PATIENT VERSUS AGENT-ORIENTED INQUIRIES	10
STRUCTURING THE DEBATE.....	11
ETHICS AND TECHNOLOGY	13
CHARACTERISATIONS OF TECHNOLOGY	15
DEFINING THE SCOPE OF TECHNOLOGICAL ENTITIES	17
INTELLIGENT TECHNOLOGIES.....	19
EXAMPLES OF INTELLIGENT TECHNOLOGIES TODAY	21
MORAL STATUS	23
MORAL CONSIDERATION AND MORAL PATIENCY	24
DIRECT AND INDIRECT MORAL STATUS	28
PROPERTY-BASED THEORIES.....	31
PROBLEMS WITH A PROPERTY-BASED APPROACH.....	35
AN ALL TOO MODERN PROBLEM	40
THE DOWNFALLS OF AN ATOMISTIC WORLDVIEW	43
RELATIONAL THEORIES.....	46
PHENOMENOLOGICAL INSIGHTS	49
COECKELBERGHS SOCIAL-RELATIONAL ACCOUNT	53
COECKELBERGHS' CONDITIONS OF POSSIBILITY	55
THE SOCIAL-RELATIONAL ACCOUNT DISCUSSED	57
THE DANGERS OF RELATIVISM.....	61
IN DEFENSE OF MORAL RELATIVISM ACCUSATIONS: PROPOSALS OF AN ETHICS OF SKILL.....	64
IMPLICATIONS OF AN ETHICS OF SKILL FOR THE MORAL STATUS OF TECHNOLOGY	65
EXPANDING THE RELATIONAL ARGUMENT	69
TECHNOLOGIES AS MEDIATORS.....	70
INTRA-RELATIONAL CONSTELLATIONS AND SYSTEMS	72
CONCLUSION	74
ENDNOTES	76

Introduction

In recent decades, the discourse on the moral status of technologies has been expanded by an intriguing new alternative. Whereas traditional ethical frameworks have mostly neglected the moral status of non-human entities due to a lack of certain properties (e.g., consciousness, intelligence, interests), scholars like Mark Coeckelbergh (2012) have put forward a relational approach – thereby seemingly challenging the anthropocentric tradition. Rather than focusing on what entities are or what they may become, proponents of a "relational turn" suggest shifting our focus toward the relations we have to said entities (Coeckelbergh, 2012; Ess, 2015; Danaher, 2020; Gunkel, 2010).

Despite offering a promising new viewpoint, I argue that both conventional property-based approaches and relational alternatives are ultimately unable to capture the moral status of novel technological entities sufficiently. Rather than offering a radically new perspective to the modernist property-based tradition, the relational approach inherits many of its same weaknesses. Just like in modernist frameworks, the social-relational argument ascribes status in a manner that too heavily relies on the perspective of the human subject – a criticism which has been similarly voiced by scholars like Tollon (2021) and Sætra (2021).

Additionally, I will point to the binary manner in which relations are understood as one of the core weaknesses in Coeckelbergh's (2012) proposal. It constitutes an outdated perspective of human-technology relations that is no longer tenable in a world where relations can no longer be understood to primarily exist between two parties but rather in an ever-growing and entangled web of moral subjects.

While the relational approach successfully recognises the moral status of some additional technological entities, it only does so from an anthropocentric perspective and fails to capture the moral character of others sufficiently.

To demonstrate this, I will refer to recent examples of technological entities developments that characterise a shift in the modern world as we know it. The accelerating pace of recent technological innovations simultaneously constitutes the motivation for my inquiry.

Applications such as artificial intelligence (AI) are not exactly new technological innovations but have been around since the middle of the 20th century (McCarthy et al., 2006). What is new about technologies today, and what poses a challenge to conventional thinking about moral status, is the intensity of their presence and adaptability. Trends like the *intelligentization* of everyday appliances and so-called *ambient technologies* have changed

the nature of our encounters with technological entities. We are no longer solely encountering technologies as physical, embodied devices (e.g., robots, computers, intelligent interfaces) but are instead increasingly surrounded by entire digitalised and personalised environments (e.g., smart sensors, doorbells, virtual assistants; Verbeek, 2011).

As technological entities have become increasingly seamlessly integrated into our everyday environments, we are at risk of becoming less and less aware of their existence –to a point at which they become almost invisible. As computer scientist Mark Weiser (1991) fittingly described in regards to computing technologies: ‘The most profound technologies are those that disappear. They weave themselves into the fabrics of everyday life until they are indistinguishable from it’(Weiser, 199, p. 66).

The ongoing transformations in the nature of technologies have prompted a shift in the world around us: Where it was once made up of stand-alone things, properties, and binary relations, the world we find ourselves in today is characterised by interactions, processes and networks (Floridi, 2015, p. 2). And yet, the moral status debate has largely failed to recognise this shift. Still, a majority of accounts question the status of technological entities as isolated subjects embodied within material objects, artefacts, machines, and so forth. As a result, it comes as no surprise that entities like virtual assistants or smart devices have been largely excluded from moral status ascriptions – both in traditionally modernist and alternative relational accounts (Coeckelbergh, 2012; Bryson, 2010; Hall, 2007; Himma, 2003).

It is with this context of a novel intelligent environment in mind that I seek to investigate and challenge the way we think about moral status. How do we go about determining the moral status of these novel artificial entities? Is it appropriate to apply the same frameworks and methods that we used to examine the moral status of other non-human entities, such as animals and nature, to the status of technological entities? What does this ultimately imply for the place of technological entities in the moral universe?

My reframing of the moral status question is broadly structured into two parts. It comprises both a reframing of how the moral status question is posed in the first place and, secondly, an investigation into how moral status ascriptions proceed.

The first part of the questioning is concerned with the question of *what* moral status is. It is dedicated to an “untangling” and structuring the current web of accounts to provide a clearer understanding of the nature of the debate and the kind of suppositions hindering its progress. More specifically, this part of the analysis aims to provide a foundational understanding of how the question of moral status has been traditionally understood. As it turns out, much of the resistance by scholars to grant moral standing to technologies can be

linked to their understanding of moral status as a matter of duties and responsibilities. Too often, the question of moral status is posed by asking which entities ought to be respected or preserved the well-being of (Bryson, 2018; Mosakas, 2020). Complicating matters further are various conflicting notions about what concepts such as “moral standing” or having “moral importance” entail in the first place. The beginning of my analysis will, therefore, be aimed at providing a better understanding of the relevant terminologies, as well as an outline of the discourse thus far. Given the sheer complexity and variety of ways the moral status question has been posed, this delimiting and outlining of accounts is necessary to define clear boundaries for my investigation.

The second part of my inquiry is a discussion and subsequent critique of the various approaches to moral status ascription. While the first part investigates *what* moral status is and how we can best understand it, my aim in this section is to challenge *how* moral status is determined and grounded. Here, the two main argumentative strategies are compared by referring to property-based and relational justifications of moral status (Coeckelbergh, 2012). While relational arguments for moral status have been proposed in a wide variety of accounts (Danaher, 2020; Floridi, 1999; Laukyte, 2019; Tavani, 2018), I will specifically focus my discussion on the social-relational framework put forward in Coeckelbergh’s (2010, 2014) and Gunkel’s accounts (2012, 2018). The comparison between these two argumentative strategies further allows me to realise a broader comparison between two vastly different ethical traditions and metaphysical worldviews.

The question of how moral status is ascribed also becomes a question of the implications our moral framework has on the ways we conceptualise entities around us. Ultimately, I will find that a traditionally Western, dualistic worldview arising from the property-based approach is untenable in the modern technological context. From a review of recent technological innovations and phenomena, it will become evident that an attempt to locate moral status within separated entities is simply not feasible anymore. Rather than locating moral status in the properties of one entity or in the relations a human agent has *to* an entity, we need to consider the *intra-relations* occurring within them. By drawing from various relational accounts (Introna, 2007; Latour, 1993), I will point to a relational view that expands beyond the dualistic paradigm and considers relations at a broader level, namely those between entire networks of technological constellations.

Scope and Objective

To clarify the objective of my own questioning, it is helpful to put it into the context of the main fields of ethics in which moral status questions typically occur in the first place. Generally, any inquiries into moral status belong to the discipline of moral philosophy, which ranges across multiple branches and ethical theories. In each of the different ethical branches, one typically finds a corresponding objective driving the scholars' inquiry.

One way of categorising amongst accounts is to distinguish between both their forms of ethics (normative, descriptive, meta- and applied ethics) and their ethical theories (utilitarian, deontological, virtue ethics) applied (Sætra & Danaher, 2022).

Accounts that fall under the branch of descriptive ethics, for example, attempt to describe how people make moral judgements and form beliefs. An example of such an account is Gray et al.'s (2012) 'mind perception theory', which argues that the moral status we ascribe to technology results from a perceived mind and perceived suffering in the entity. Descriptive theories are helpful in providing insights about how we make moral judgements and why some technologies are more likely to be perceived as moral patients than others. However, they do not provide an answer to what is morally "right" or "wrong", as this would be the task of normative ethics (Copp, 2005).

The majority of accounts discussing the moral status of technology turn out to belong to the branch of normative ethics. The question of moral patiency in robots, for example, is commonly posed to provide normative answers on how we should act, what obligations we have, or whether a robot can be morally wronged in the first place (Floridi, 1999; Tollon, 2020; Torrance, 2013; Umbrello & Sorgner, 2019).

Gunkel (2012, p. 6), for instance, asks 'whether and to what extent robots, machines, and so on might constitute an other to which or to whom one would have appropriate moral duties and responsibilities to'. In other articles, his question simply is: 'Can and should robots have rights?' (Gunkel, 2018). Questionings of this kind have prompted the development of a number of specialised fields, such as robot ethics (Lin et al., 2011; Tzafestas, 2016) or machine ethics (Anderson & Anderson, 2011). Such sub-fields of moral inquiries are specifically concerned with questions of moral duties and responsibilities. They also often overlap with questions of applied ethics – a field which additionally sets out to develop concrete frameworks and guidelines for our behaviour (e.g., within engineering or computer ethics; Copp, 2005; Sætra & Fosch-Villaronga, 2021). This is reflected in familiar concerns such as: 'Can we attribute responsibility to machines if they make mistakes?' or 'Can

algorithms or robots be designed to act morally?’ (Copp, 2005). The aim of posing the moral status in applied ethics is to draw practical consequences for everyday applications and does not reflect the objective of my own questioning.

Similarly, the question about whether an entity has rights or deserves a specific kind of moral consideration is one of normative nature and is not pursued in my analysis. The motivation of my inquiry is not to answer “how we should act” towards technologies or to provide normative reasons for doing so. My aim is neither to identify which of our actions towards technological entities can be considered right or wrong or what kinds of moral obligations we have towards them. Nor do I intend to draw conclusions for applied ethics or make practical recommendations for our interactions with technology. Instead, I seek to provide a broader analysis of the nature of the moral status question in the case of technological entities. The primary focus of my inquiry is determining how the moral status of modern technologies can best be understood and approached, and less on what the exact moral status of an entity is. It is somewhat of a “questioning of the questioning” to investigate the ways in which the question has been conventionally approached in the first place. The sought-after insight is not to determine *how much* we should consider technological entities – this will later be defined as a question of *moral significance* rather than moral status – but rather *what* we should consider about them. Should it be their properties and ontological conditions? Or is it the social situatedness and relationships within entities that should be considered for their moral status?

Although the goal of my endeavour is not to provide specific normative answers, it is not necessarily irrelevant for practical implications and insights. Investigating the moral status debate in this way will show that how we approach questions of moral status *does* matter for how we think about our daily, lived engagement with technologies. In the course of my discussion, it becomes clear that the ways in which moral patiency is understood inevitably impact the moral relevance we ascribe to our daily actions and, therefore, offer crucial insights for normative questions. How we think about moral status and how we go about determining it in technological entities directly impacts the kind of awareness we have about their moral significance and role in social situations. How we should act as a result of this heightened awareness, however, will remain outside the scope of my analysis. As Paul Verbeek (2011) stated in reference to Gerard De Vries, the point of departure for moral inquiries should not be a question of ‘how should we act’ but rather a question of ‘how to live’ (De Vries, 1999). If there is any practical applicability of my analysis, then it is to the

ways we choose to live and exist in a world full of technological entanglements and relationships.

Hoping to have made my objective clear at this point, I will now move to define the scope of my investigation in the following sections. The scope of my investigation will be confined both by the (1) specific kinds of technological entities being questioned, (2) the definition of moral status applied, and lastly (3) the strategy through which it is ascribed. The latter requires a more in-depth analysis of the main argumentative strategies taken and constitutes the second part of my inquiry.

To define the scope of my questioning more precisely, I will first begin with a review of the moral status debate thus far. From this, the vast variety and complexity of the debate will quickly be noticeable, which prompts me to specify the bounds of my inquiry further. Within this analysis, only entities that qualify as *intelligent technologies* and so-called *patient-oriented* inquiries will be relevant. The rationale for setting this focus will emerge following the overview of the current discussion as a whole.

The Moral Status Debate Thus Far

Within the domain of technology ethics, the question of “What matters morally?” has been posed in numerous ways and has resulted in just as many contrasting answers. Attitudes from scholars range from a complete denial of moral status or the importance of technological entities to proposals of machine rights and full moral status in AI. On one side are scholars strongly opposing the idea of moral status in technological entities – some even find the discussion trivial (Birhane & van Dijk, 2020). One such example is Bryson’s (2010) article ‘Robots should be slaves’, in which he claims that robots cannot and should not be morally considered. On the other end are scholars arguing in favour of moral consideration of technological entities (Davies, 2011; Levy, 2009)– some even going as far as finding them to have full moral agency (Floridi, 2004). Others like James Moor (2006) take a more nuanced view and propose to ascribe moral status across varying degrees: Certain kinds of technologies like robot jockeys would have some moral importance as ‘ethical impact agents’, yet they would not be granted the status of a ‘full ethical agent’ (Moor, 2006). The variety of answers given by scholars arises from entirely different notions and assumptions about moral status, which I will specify first.

Patient Versus Agent-Oriented Inquiries

For a simplified overview of the debate, I will broadly distinguish accounts between the two primary forms of moral status questioned. Although some have suggested attributing moral status to various degrees and types (Moor, 2006; Wetlesen, 1999), I will here only differentiate moral status into two main variations: *moral agents* and *moral patients*.

The intuition behind the patient/agent classification is reflected in a dyadic understanding of moral status and is similarly found in the concept of the moral dyad (Schein & Gray, 2018). The basic idea of the moral dyad holds that any moral situation fundamentally consists of at least two parties. How these two parties are referred to varies amongst scholars, yet one of the most common articulations is to speak of patients and agents (Floridi & Sanders, 2004).

While definitions of the terms vary, the most general understanding is that moral patients are entities that can be acted upon, whereas moral agents are entities from which said actions originate from (Floridi, 2004). Applying this terminology to the moral status debate allows us to distinguish accounts as being either *patient-* or *agent-oriented* or *standard* and *non-standard* (Floridi & Sanders, 2004; Gunkel, 2012). Similar distinctions have been formulated by several scholars (McPherson, 1984; Miller, 1994; Regan, 1983); however, I will primarily apply Floridi and Sanders' (2004) characterisation here.

The first, standard position, represents the most popular view of moral status throughout the history of ethics (Floridi & Sanders, 2004). A standard view holds that only entities that are moral agents can qualify for the role of a moral patient (Floridi & Sanders, 2004). In an agent-oriented perspective, moral status is questioned in its highest degree of a moral agent or as *full moral status* (FMS). Within technology ethics, accounts inquiring from this perspective are typically motivated by concerns about machine responsibility, autonomy, and accountability and ask whether machines can be held responsible for their actions (Bigman et al., 2019; Chinen, 2016; Tigard, 2021).

In the alternative view, scholars pose the question of moral status as moral patiency, sometimes also phrased as questions of 'moral standing' (O'Neil, 1997) or 'moral considerability' (Goodpaster, 1978). A quick clarification of my own use of the terminologies in the following: I will apply the term "moral status" to mean entities that qualify as moral patients and specifically use "full moral status" to mean entities qualifying for both agency and patiency. Alternative terminologies such as "moral standing", "moral significance", or "moral consideration" require further context and will be discussed in a later section.

A non-standard approach holds that all moral agents automatically qualify as moral patients. However, entities that lack moral agency are not necessarily excluded from moral patiency or moral status (Floridi & Sanders, 2004). This implies, for example, that a robot can be considered to be a moral patient without being given the status of an agent. Inquiries of this kind have gained popularity following post-environmentalist efforts to ascribe the moral status of a lesser kind (O’Neil, 1997; Taylor, 1986; Wetlesen, 1999). The aim was to include non-human entities in the moral circle and to counter the frequently criticised “human supremacism” across moral philosophy (Estrada, 2020). By now, the idea of entities having moral status other than FMS is widely accepted and barely appears to be a “non-standard” view anymore. The increased interest of scholars in patient-oriented perspectives is clearly noticeable in the discourse on technology: More and more of the discourse is concerned with the status of artificial entities to determine whether they should be given moral consideration or not. This is especially evident within computer or robot ethics, where one of the most common questions centres around our moral obligations and duties towards technological entities (Harris & Anthis, 2021).

In the following discussion, precisely these patient-oriented inquiries will be investigated. Accounts questioning moral status in the traditional sense of an autonomous, independent, and intentional agent (Johnson, 2006) will not be relevant and would require a separate discussion.

Structuring the Debate

Although limiting the debate to a patient-oriented perspective allows to narrow the inquiry to a smaller subset of accounts, the discourse still remains highly complex: The recent literature contains a vast variety of contrasting and overlapping discussions, often caused by differences in the scope of technologies questioned and conflicting underlying assumptions made. Some question technologies in all their possible variations, including artefacts and tools (Frederik et al., 2011), whereas others specifically inquire about computer systems (Johnson, 1985; Moor, 1985) or social robots (Coeckelbergh, 2012; Gunkel, 2012; Verrugio & Operto, 2006).

Further adding to the complexity of the debate is an inconsistent use of terminologies and varying notions about what it means to be a moral patient or to give moral consideration in the first place. In order to sufficiently identify the underlying assumptions steering the debate, I will discuss accounts in two subsequent aspects: How moral status is defined and understood and, secondly, how moral status is determined and attributed.

The first is a matter of definition and understanding. A comparison between accounts is increasingly complicated due to a wide variety of terminologies applied by scholars: Some refer to moral patients as those who have ‘moral considerability’ (Goodpaster, 1978; Hale, 2011) and belong to the ‘moral circle’ (Danaher, 2020), whereas others speak of ‘moral importance’ (Johnson, 2006), or ask about ‘moral standing’ (O’Neil, 1997). Especially the term “moral consideration” consistently reappears in connection to moral status. Yet again, scholars often lack a shared understanding of the concept as well as its relation to moral status: Is granting moral consideration the same as ascribing moral patiency? Or are the two understood as separate notions from one another?

Similar questions arise in regard to the term moral status itself. Some scholars like Floridi (2004) or Benjamin Hale (2011) have suggested a more controversial understanding, which conceptualises moral status as entirely separate from moral standing or moral consideration. Having moral status would, therefore, not automatically warrant moral consideration and vice versa (Hale, 2011). The majority of accounts, however, apply the terms of moral consideration and moral status synonymously (Floridi & Sanders, 2004). The specific applications of these terms will later be expanded upon and specified.

Following this first differentiation between terminologies and understandings, we can further distinguish accounts in the way moral status is determined. As mentioned earlier, the characterisation applied in this paper is based on Gunkel’s (2012) and Coeckelbergh’s (2012) distinction between property-based and relational strategies. An alternative characterisation is to differentiate between so-called *direct* and *indirect* justifications of moral status (Samuelsson, 2008). However, applying these terms only complicates matters further as the meaning of the “direct” vs “indirect” argument remains ill-defined and debated within the literature. Relational accounts, for example, are generally regarded as making an indirect argument and are criticised for such (Sætra, 2021). Yet, it can also be argued that a relational argument is not necessarily indirect and allows for a direct justification of moral status the same way a property-based argument does. I will later expand on this point and argue for an understanding of direct moral status that does not rely only on intrinsic or extrinsic properties alone.

An interesting aspect to note is that the kind of argumentative strategy scholars apply (property-based or relational) often directly corresponds to specific ethical frameworks. A property-based argument would generally correspond to the dominant ethical theories of Western modernity, mainly consequentialist (e.g., utilitarianism) or deontological (e.g., Kant’s “Pflichtethik”, 1785). A preference utilitarian argument made by Peter Singer (2003),

for example, defines a morally right act as one that maximises the fulfilment of preferences of entities (Singer, 2003, p. 257). Naturally, such an argument implies that only entities with the capacity to have preferences can be morally relevant. For the moral status of technology, a utilitarian perspective has typically resulted in the exclusion of technological entities, as they were considered to lack feelings (or interests, in other kinds of utilitarian theories) that could be harmed or benefited. Scholars like Lender (2016), who apply a utilitarian framework to the moral considerability of AI, argue that moral status can only be ascribed if an entity has phenomenal consciousness and, thereby, interests that must be accounted for.

On the contrary, frameworks of virtue ethics are commonly associated with *indirect* justifications of moral status and relational accounts. The basic idea underlying many such arguments is that our behaviour towards technology can cultivate either vices or virtues. Therefore, moral consideration towards other entities is encouraged to foster good character and, as a result, virtuous behaviours towards other humans (Coeckelbergh, 2012; Hacker-Wright, 2007; Sparrow, 2020). In the following, however, I will not subscribe to one specific ethical framework to define what a morally “good” act is and will consider various understandings.

Overall, the moral status discourse on technology is clearly much more extensive in depth and scope, as it can be addressed here and requires some further clarifications. Given the variety of overlapping discussions across accounts, I will first address some of the emerging sub-questions by asking:

- (1) Which kinds of technologies are inquired about?
- (2) Which definition and understanding of moral status is applied?

Following these clarifications, the main question becomes one about how we can best think about the moral status of these entities in the first place. Specifically, which approach and framework is most suitable for how we think about the status of technological entities?

Ethics and Technology

To specify which kinds of technologies are inquired about, it needs to be clear what it means when we speak of “technology” in the first place.

Within the context of moral status, the term technology has been primarily used within domain-specific inquiries ranging from computer ethics (Moor, 1985), robot ethics (Gutman et al., 2012), machine ethics (Wallach et al., 2008), to nanoethics (Allhoff et al.,

2007). While each of these fields denotes a specific focus of interest, they can all be said to broadly fall into the discipline of *philosophy of technology* (Ellul, 1964; Mitcham, 1994).

The rapid growth of interest in the field can undisputedly be attributed to the social and technological developments fueled by the emergence of information and communication technologies (ICTs) and AI over the last two decades. Yet, ethical concerns about technology have been subject to moral discussions all throughout history and have been around as early as the ancient Greek period. During this time, technology was more commonly thought of as *téchne* and hereby conceptualised as craft, art, and practical skill (Skrbina, 2014). Long before Science fiction fostered fears of a life-threatening superintelligence, Plato's 'Phaedrus' notoriously criticised technology in the form of writing and cautioned readers about the dangers of the tool (Churchill, 2003).

Throughout literature, we can find various definitions of the term technology itself: While some philosophers derive their understanding of technology from the Greek origins of the term, definitions of technology throughout the 20th century range from 'technology as applied science' (Bunge, 1974) to technology as 'human making or using of material artefacts in all forms and aspects' (Mitcham, 1978, p. 232) to broadly as a 'reshaping of the physical world for human purposes' (Rapp, 1989).

Further ambiguities about the meaning of technology arise due to a wide range of possible understandings: We can speak of technology in the sense of active-making and skill, thereby referring to a 'domain of knowledge and processes' (Bigelow, 1829) and a 'domain of human activity' (Ropohl, 1990). In other contexts, technology is applied to describe a specific object or material entity, thereby referring to a 'domain of objects' (such as artefacts, tools, and machines; Reydon, n.d.).

Today, some of the most impactful technologies can hardly be considered material manifestations and would no longer be allocated within a domain of objects. Instead, technology has transcended physical and geographical borders and now forms entire integrated networks and systems of entities. Examples include ICTs, WiFi communication infrastructures, cloud infrastructures, computing software, algorithms and more.

To reflect this modern development, I will apply the term "technology" based on a distinction proposed in Deborah Johnson's account, which differentiates between *artefacts* and technology (Johnson, 2006, p. 196). Here, 'artefacts' refer to material and static objects and are not the entities primarily focused on in my inquiry (Johnson, 2006, p. 196). Instead, I intend to apply the term technology to mean a combination of artefacts, social practices, and systems of knowledge, making up entire 'socio-technical ensembles', 'socio-technical

networks' and 'systems' (Johnson, 2006, p. 196). Defining technology this way allows for an understanding beyond mere objects and reflects a view of technology as hyperconnected networks of entities within a specific sociocultural context.

Characterisations of Technology

While it is not my interest to discuss the nature of technologies and whether they are morally “good” or “bad”, it is worth noting that the outcome of moral status discussions is often closely related to the scholar’s characterisation of technology. Within the moral status debate, accounts can broadly be allocated into two main perspectives on technological entities: *Instrumentalist* and *deterministic* views. To this day, the most widely held definitions of technology stem from an instrumentalist viewpoint.

Instrumentalism reflects the fundamental understanding of technology as a tool and is famously held by scholars such as Andrew Feenberg (1991), John Haugeland (1981), Daniel C. Dennett (1996), and partly Martin Heidegger (1993).ⁱ Heidegger (1993) describes the nature of technology as an instrument serving human purposes: ‘Technology itself is a contrivance – in Latin *instrumentum*’ (Heidegger, 1993). This is similarly expressed by computer scientists such as John Searle: ‘The computer is a useful tool, nothing more nor less’ (Searle, 1980, p. 190).

The instrumental characterisation appears strikingly fitting towards objects and utensils, which have been used as tools and extensions of human abilities all throughout history. As Marshall McLuhan famously characterised, technologies are media through which we are able to extend our natural or bodily abilities (McLuhan et al., 1995). As technology is merely an extension of men, the instrumentalist view holds that technology is free from having an interest or purpose of its own. Any purpose is only derived from the ends of the human user. Thus, scholars commonly characterise technologies as morally neutral and with no valuative content of their own (Feenberg, 1991).

However, as more discussions on machines and artificial intelligence (AI) arise, the tool characterisation of technology has become increasingly challenged. An AI system not only seems to be an extension of our bodily functions but also appears to exist independently from the user – or even as autonomous. The idea of technology as an individual source of power is reflected in accounts applying a deterministic view of technology (Winner, 1977). Within a determinist theory, technology is considered an autonomous force beyond human control and is ‘value-laden’ instead of neutral (Winner, 1977).

In a time of intelligent technologies and artificial intelligence, the instrumentalist view of technology is increasingly challenged. Characterisations like Langdon Winner's description of value-laden artefacts (1977) appear increasingly accurate when considering phenomena such as algorithmic biases. Many Actor-network theorists who characterise everyday objects as 'political locations' (Introna, 2007; Latour, 1993; Latour, 2005) express a similar idea of technology as locations of moral value ascription.

Whether the nature of technology is that of an instrument or not is not necessarily relevant to the question of its moral status. How we should conceive the phenomenon of technology *per se* would require an entirely separate analysis of its character, as it has been undertaken in numerous other accounts (Feenberg, 2003; Idhe, 1993; Rapp, 1989; Winner, 1977). The character of technology entities and their moral standing are clearly two separate matters of inquiry. And yet, accounts like Idhe's (1993) and Winner's (1977) and their famous characterisations of technology continuously reappear in moral status debates. The way technologies are characterised is undeniably tied to our perception of their moral status:

Within recent literature, it is noticeable that the notions scholars have about the nature of technology and the notions they have about its moral status are correlated. How the nature of technology is conceptualised is often directly related to argumentation in favour or against moral status ascription. Scholars opposing the idea of moral agency in technology usually do so from the instrumentalist perspective, which views technology as merely an extension of a human's interest – therefore incapable of having responsibility on its own (Goertzel, 2002; Haugeland, 1981; Lyotard, 1984). Similarly, those who exclude technological entities from moral patiency do so by arguing that they do not have a "good of their own" or "intrinsic interests" that need to be considered (Dracopoulou, 2003). Because technology is viewed as an instrument, its interests are only "derivative" from their human creators (Taylor, 1986).

Even if we find the instrumentalist view of technology to be correct, it does not have to stand in the way of moral status ascriptions. Neither does it have to conflict with a non-neutral characterisation of technology: One understanding of technology should not mean that we have to exclude the possibility of another. Unfortunately, however, a majority of the moral status debate associates the instrumental view of technology with a lack of its moral status. Throughout my analysis, I hope to challenge the supposition that an entity's characteristics are indicative of its ability to gain moral standing. By reconsidering how we think of moral status, it is possible to acknowledge that technology is both an instrument for human purposes and inherently value-laden and "morally charged" at the same time. While a complete characterisation of technology will not be pursued, I will introduce some concrete

examples of artificial entities in the following, which challenge the neutral view of technology. This will provide further reason to reconsider the traditional ways in which the moral status of these entities has been approached.

Defining the Scope of Technological Entities

Within the moral status debate, the ever-increasing range of technological entities has led scholars to address technological entities in dozens of varying scopes and manifestations: Capturing the broadest range of entities are scholars who discuss technology in the form of all artefacts: From chairs to microscopes to nanochips (Frederik et al., 2011). Other accounts take a more specified approach and limit the discussion to Information and Communication Technologies (ICTs; Floridi, 2007; Ganascia, 2015) and to specific entities such as ‘social robots’ (Coeckelbergh, 2018; Gunkel, 2018), algorithms (Jaynes, 2021) or ‘artificial life’ (Sullins, 2005).

Resulting from this is a multitude of domain-specific inquiries in which each technological entity is questioned individually. Within increasingly narrow sub-domains, scholars are separately questioning ‘the moral status of AI’, ‘the moral status of driverless cars’, ‘the moral status of robots’, and so forth (Sætra & Danaher, 2022).

In this thesis, I choose to avoid setting a narrow focus of this kind: Defining the scope of technological entities too tight can ultimately hinder progress in the moral status debate and is problematic for several reasons. To illustrate this, let’s consider the field of robot ethics as an example. According to a recent literature review, robots are currently the most frequently questioned candidate for moral status (Harris & Anthis, 2021). In several accounts, the inquiry is even more specifically limited to robots in humanoid form or social robots (Brey & Søraker, 2009; Gunkel, 2018; Robertson, 2014).

When the moral status question is explicitly posed towards social robots, the discussion tends to centre around the robot’s unique properties, their ability to feel or act autonomously, the future potential for sentience, or their role within society. The resulting arguments made by scholars are then often specific to the individual characteristics of the entity itself. In J. Danaher’s ethical behaviourism, for example, he argues that moral status should be ascribed to robots if their observable performance is equivalent to humans (Danaher, 2020). Others similarly ground their ascription of moral status in humanoid appearances and our human capacity to empathise and care for robots (Dixon, 2015; Schmetkamp, 2020). Arguments of this kind are naturally not transferrable to other robotic manifestations, such as drones or aerospace robots, but only to those who meet the criteria. If

the entire discussion of moral status in technologies were to proceed this way, every newly developed category of technologies would have to be approached individually and would require its own set of arguments.

Considering the fast pace of technological progress, arguments based on the current characteristics of an entity are already sub-optimal in that they cannot account for future developments. They only apply for as long as we don't witness any fundamental technological changes within the digital world. The bigger problem, however, is that questioning the moral status of each entity separately poses the risk of focusing too much on its unique characteristics (e.g., its specific properties or the nature of our relation to it), whilst some of the shared fundamental truths and phenomena across different technological entities are overlooked.

Questioning technologies within their specific sub-domain can lead to all too well-known discussions of differences: What makes humanoid robots deserving of moral status and not autonomous cleaning robots? Justifying why one domain of entities deserves moral status as opposed to the others based on its features has proven counterproductive to any progress: It requires a solid justification of why one characteristic possessed by an entity takes precedence over another. This becomes especially evident in a review of property-based arguments later on. Meanwhile, some of the shared underlying insights and ethical implications across novel technologies as a whole remain overlooked. Inquiries that only focus on social robots with humanoid appearances, for example, are too narrow in that the potential artificial suffering of an extended category of entities (i.e., softbots) is neglected. This is potentially worrying, given that the majority of artificial entities that will likely have the most significant impact on the future of humanity do not appear in the embodied or humanoid form at all – even remaining entirely hidden from our imminent perception.

In order to capture the variety of technological entities at a wide enough level while still emphasising their modern manifestations, I will focus my inquiry on so-called 'intelligent technologies' (Brynjolfsson & McAfee, 2014). The word "intelligence" is known to spark controversies; however, in this case, intelligent technologies are not meant to denote the property of an entity. As Molina (2022) has pointed out, the term should not be understood as 'a rigid characterisation to determine whether a system is intelligent or not' (Molina, 2022). Within recent literature, 'intelligent technologies' is commonly applied as an overarching term for the latest generation of novel technologies, ranging from AI systems, neural networks, and automated computing to additive manufacturing (Marinov, 2019; Venturini, 2022).

In contrast to less-novel technologies and mere artefacts, intelligent technologies differentiate themselves in their ability to learn from experience, adapt to new data points, and gather, analyse and respond to the data of their surrounding environment (Marinov, 2019; Venturini, 2022).

In this way, they most closely resemble concepts such as those of ‘intelligent agents’ (Zhong, 2001) or ‘multi-agent technologies’ (Cointe et al., 2016) within the field of AI. The most widely accepted definition of an intelligent agent is that of a computer system situated in some environment, and that is capable of flexible, autonomous action in this environment in order to meet its design objectives’ (Wooldridge & Jennings, 1994). More specifically, I apply the term ‘intelligent technologies’ to reflect the so-called *weak notion* of intelligent agents. While specifications vary, the most commonly agreed-on features of a weak agent are flexibility, interactivity, and autonomy. In contrast, a strong notion of intelligent agents requires additional attributes such as mental states (Weiss, 1999). An example of such entities are hardware devices like smartwatches or sensors, which perceive their environment – in this case, the physical world or the human user through an interface – and respond to changes in it autonomously. My application of the term ‘autonomous’ does not denote some human-like mental property but merely the ability of a device to operate without the need for direct intervention of a human user (Castelfranchi, 1995).

Intelligent Technologies

For the purpose of my inquiry, the term ‘intelligent technologies’ is intentionally applied rather loosely to comprise a greater variety of software and hardware applications within the domain of Artificial Intelligence programs. The applications of AI used today range in varying complexity from so-called *narrow AI*, designed to perform singular tasks, to *strong AI* and visions of *artificial general intelligence* (AGI), aimed to replicate human intelligence. Examples of weak AI applications are social media algorithms, which use machine learning algorithms to tailor suggested content to the human user’s specific interests or behaviours (Kaiser et al., 2020).

In our daily lives, we constantly encounter AI systems in various forms and manifestations, often without even being aware of their presence. In the digital sphere, intelligent technologies such as machine learning algorithms are present in every Google search output, every filtered email inbox and across almost every Online Shopping site we visit. Outside of the virtual, we find intelligent technologies in their material manifestations, from smart watches collecting health data to collaborative robots assisting factory workers.

With the help of the Internet of Things (IoT), many of these devices today are able to communicate with each other within interconnected networks. In contrast to AI programs, which do not necessarily require interconnectivity over the web, the IoT comprises physical devices directly connected over the internet, thereby forming the entire global infrastructure. One example of a practical implementation of IoT devices are *smart homes* in which anything from door locks, thermostats, security sensors, or robot vacuum cleaners can be interconnected and responsive to inputs in the environment. A smart fridge, for example, would have an RFID (Radio Frequency Detection) tag embedded in every shelf to scan which grocery items are put inside. It can then share this information directly to the smartphone of the human user to let them know what items they are short on or to make healthy meal recommendations (Rouillard, 2012).

One way to describe these kinds of technologies, that appears to be especially fitting for these examples, was put forward by Elena Esposito (2022): In her book on algorithms, she encourages us to speak of intelligent technologies in terms of *artificial communication* devices rather than *artificial intelligence*. It is precisely the element of communication and interaction that I hope to emphasise throughout my analysis. The reason today's technologies are so much more impactful is not that they have become smarter but that they have become better at communicating with us (Esposito, 2022).

For now, we still have a number of "lifeless" artefacts around us: Furniture items, clothing, or utensils are, as Floridi (2007) describes, 'non-interactive, irresponsive and incapable of communicating, learning or memorising' (p.7). But with the onset of AI and IoT systems, even the most ordinary objects are likely to become more and more interactive: Pen and paper are replaced with iPads and supplemented with automated writing suggestions; hospital beds become intelligent and track our health data; and various other artefacts are integrated into the IoT. The key feature here is that intelligent technologies are malleable in their nature. As intelligent devices gather and respond to our inputs and data, the morality of our actions becomes quite literally imprinted within them. As a result, our behaviours and actions translate into real-time impact – once again making the question of their moral standing ever so urgent.

The full scope and variety of intelligent entities and applications cannot be done justice at this point and would require a much more extensive discussion. What matters is that posing the moral status question at this broader scope allows for a different view of technologies in systems. Understanding the moral status of intelligent technologies is hardly possible by analysing them as isolated entities. To investigate their morality in full depth, it is

crucial to acknowledge how they are situated in our world – namely in interdependencies with other entities or networks.

Inquiring about the moral status of intelligent technologies overall naturally encourages a different view of entities within systems rather than stand-alone subjects. I argue that this is entirely appropriate and necessary, given the scope and amplitude of our moral actions today. The recipient of our actions today can rarely be confined to a single entity or counterpart anymore, especially when actions are carried out in the virtual sphere.

It further emphasises the novel nature of human-technology interactions today – which then brings about other questions: Are technological entities today really so drastically different that we need to think about their moral status in a new way? Why should modern technological phenomena and interactions even prompt us to reconsider their moral status? By comparing various moral arguments and applying them to exactly these intelligent technologies, I aim to make the answer to these questions self-evident. Additionally, I hope to reveal some of the challenges that both property-based and relational arguments face when it comes to more “unusual” candidates for moral status.

Examples of Intelligent Technologies Today

One specific example that I will introduce to emphasise the individualised and interconnected nature of intelligent technologies even further is that of search algorithms. The most powerful search engines today, such as Google or Bing, implement AI in every step of the process: Natural language processing (NLP) is used to better understand our written language and identify user intent behind a search query; Machine Learning algorithms learn and analyse our interactions and behaviour to identify the best possible search outputs, filter keywords, generate search suggestions and so forth (Nick, 2021). The result of every Google Search is the product of complex ranking algorithms, taking into account both overall trends amongst all users as well as individualised factors.

The factor of individualisation is especially relevant for how we think of our relations to these entities: Not only do algorithms learn and adapt to our past behaviour (e.g., time spent on specific web pages or click-through rates), but they also take highly personal factors into account – from age group, income class, political affiliation to geographic location. If someone logs in from San Francisco one day and from New York on another, Google will already register the user as a bicoastal traveller and make search suggestions accordingly. Even the type of browser being used and the kind of smartphone that a request is coming from are used by Google to make guesses about economic status and interests (Pariser, 2011).

Features like these are, of course, not unique to Google or other search engines. Personalisation is just as much at work in Social Media applications, where every like, comment and interaction is stored to create targeted advertisements and individualised timelines.

The implications are problematic in any case: In her book *The Filter Bubble*, Eli Pariser (2011) describes a sort of informational determinism in which everything we have clicked on in the past determines the content we see in the future. As a result of this, we become stuck in our very own filter bubbles: Existing ideas and beliefs are continuously being strengthened and reinforced by the output of search engines, as this will generate the highest level of interaction and satisfaction from the user. Meanwhile, the exposure to opposing belief systems and perspectives is minimised. Additionally, the kind of content and search outputs that are most likely to be promoted are simultaneously the most likely to be detrimental to the moral character of society.

According to sociologist Danah Boyd (2009), the cause for this is deeply rooted in our human nature and cognition:

Our bodies are programmed to consume fat and sugars because they're rare in nature...In the same way, we're biologically programmed to be attentive to things that stimulate content that is gross, violent, or sexual and that which is humiliating, embarrassing, or offensive. If we're not careful, we're going to develop the psychological equivalent of obesity. We'll find ourselves consuming content that is least beneficial for ourselves or society as a whole.

From a perspective of virtue ethics, this could imply morally devastating consequences. Filter bubbles and search algorithms are able to foster virtues or vices within society at a never-before-seen speed. Social media posts that exhibit virtues like humility are doomed to attract less engagement and less exposure than those which showcase extremities or even violence (Carretié et al., 2001). On the other hand, vices like 'self-indulgence' and 'vulgarity', as described in Aristotle's *Nicomachean Ethics* (1955, p. 104), would become more prevalent.ⁱⁱ

How deeply problematic algorithmic biases and filtered search engine results can become has been pointed out by numerous scholars by now: Safiya Noble (2018), for example, speaks of 'Algorithms of Oppression' to describe how data discrimination and algorithmic bias have reinforced racial stereotypes and social tensions. Other evidence of bias within Google's search algorithms was presented in a study by Baker and Potts (2013). Here, Google's autocomplete function was found to make predictions based on certain stereotypes about minority groups, which have been identified to be sexist, racist, and homophobic at

times. Effects like these even occur at a larger socioeconomic level and have the potential to influence political dynamics through so-called Search Engine Manipulation Effects (SEME) (Epstein et al., 2018). One example here are search results of political candidates, in which negative search terms are being artificially suppressed or altered to shift voting preferences (Epstein et al., 2018). Technology does not only appear to be capable of shaping our own selfhood and character but also shapes the character of entire societies.

Algorithms and search suggestions not only influence how we perceive the world around us but also determine *what* we perceive and know. In some cases, we find that certain perspectives and pieces of information become entirely inaccessible due to the process of data discrimination (Noble, 2018). Certain data or webpages simply remain entirely unavailable to us. Google itself admitted to manually blacklisting certain webpages or modifying search results while meanwhile promoting other websites out of private interests (Metz, 2011).

Examples like this not only pose great ethical concerns but also challenge how technologies are traditionally conceptualised and understood. The ability of search engines and algorithms to sway opinions and foster societal values illustrates just how radically the role of technology has changed. An instrumentalist characterisation alone no longer appears sufficient in the case of algorithms. They clearly prove themselves to have moral relevance, yet the question remains whether this matters for an attribution of their moral status. To answer this, I will move to an overview of the key terms and definitions surrounding the term moral status to better establish what we mean by it in the first place.

Moral Status

Moral status is in no way a straightforward concept. The terminologies that are currently applied to the question of moral status range from *moral considerability*, *moral patiency* and *moral relevance* to *moral significance* – often with inconsistent notions of what each term implies. In the following, I will review some of the most prominently held definitions and understandings of the terminology and their applications throughout accounts. Hereby, some reoccurring themes and commonalities amongst accounts become visible. Certain understandings turn out to be inherently more popular than others, yet they are not necessarily more suitable. Specifically relevant and in need of further clarification are the notions of direct and indirect moral status, as well as the idea of moral consideration in connection to moral status.

Moral Consideration and Moral Patiency

When moral status is discussed, the default assumption is that having moral status requires the moral consideration of an entity in some way or form. Gunkel, for example, assumes that not considering an entity would exclude it from any ethical concerns and would situate it outside the ‘moral circle’ (Gunkel, 2012, p. 161). The question of moral status is predominantly asked as a question about whether or not we have a moral obligation or duty to consider an entity. Warren (1997, p. 3), for example, defines moral status as being ‘morally considerable, or to have moral standing. It is to be an entity towards which moral agents have, or can have moral obligations.’

The term “moral consideration” was first coined by Kenneth Goodpasters (1978) in his essay *On being morally considerable* in an attempt to provide clarity about the framework of moral consideration. His definition asks about the necessary and sufficient conditions for an entity X under which the following is fulfilled: ‘For all A, X deserves moral consideration from A. A ranges over rational moral agents’ (Goodpaster, 1978, p. 309).

The question of what makes an entity morally considerable, however, did not emerge with Goodpaster’s (1978) essay but was posed by several scholars prior under various terminologies. G. J Warnock (1971, p. 148), for example, applies the term ‘moral relevance’ to ask what the conditions of being considered by agents are. Others, like Christopher D. Stone (1972), use ‘moral standing’ to ask how we determine who or what may be relevant to our moral decisions.

What exactly the act of giving moral consideration entails remains contested and unclear among scholars. Goodpaster (1978) defined consideration as something that ‘broadly includes the most basic forms of practical respect’ but also noted that this may or may not include the possession of rights by a morally considered entity. Others, like Mark H. Bernstein (1998), described moral considerability as ‘a capacity to absorb moral consideration’. However, this capacity remains vaguely defined and implies a problematic view of moral consideration as some property-like capacity that is found in some entities and lacking in others.

While the question of moral agency will not be discussed in this thesis, taking a brief look at the conventional understandings of agency allows us to draw an analogy to the way in which moral patiency is conventionally understood. The common characterisation of moral agency and what Himma calls the ‘standard view’ (2009, p. 19) understand agents as being capable of some intention that is motivating their actions. Similarly, Daniel Dennett describes

agents as being intentional systems that are capable of mental states (such as beliefs or desires) which motivate an intentional act (Dennett, 1998). This way of thinking about agency shares the central doctrine that action requires intentionality – a doctrine that, within contemporary philosophy, is often contributed to the works of Anscombe (1957) and Davidson (1963). Breathing, for example, would not be considered an action as it does not require intention, and in Himma's (2009, p. 19) terms, it would simply be a 'doing'.

Some philosophers suggest even narrower criteria for agency in addition to intentionality, such as the ability to understand and act on moral principles and differentiate between right or wrong or the capacity to act out of free will (Moor, 2006). Regardless of which standard criteria of agency apply, the common underlying assumption is that being an agent requires at least some level of autonomy and intentionality, which then implies responsibility for one's actions. As Himma pointed out, the standard notion of moral agency ultimately aims to define which entities' behaviour is subject to moral requirements and can thus be attributed responsibility to' (2009, p. 21).

This has resulted in agency being traditionally ascribed only to adult humans as the only entity capable of rational thought and reason (Descartes, 1637). Others, like animals or infants, were excluded from agency and not considered to be targets of blame for their actions. Defining only human adults as agents seems intuitively correct. For example, an action done by a newborn infant (e.g., spilling food on an important document) wouldn't be judged as a moral or immoral act in a daily context. Although the action possibly causes distress and suffering in a human agent (e.g., the parent), it lacks intentionality, and the infant would ordinarily not be blamed for having done something immoral.

This intuitive understanding is typically applied by scholars examining agency in artificial entities as well: In Gunkel's (2012) investigation of 'the machine question', the question of moral agency is posed 'by asking whether and to what extent machines might be considered a legitimate moral agent that could be held responsible and accountable for decisions and actions' (Gunkel, 2007, p. 6). This view of agency is echoed by several other scholars: Mikko Siponen, for example, argues that by granting moral agency to artificial systems, we would 'start blaming computers for our mistakes' (Siponen, 2004, p. 286). Further, Hall (2007, p. 349) points out that 'moral agency breaks down into two parts: rights and responsibilities'. In his conception, an animal or a machine would not be a moral agent as it can not be held responsible for its actions and so 'morality rests on human shoulders' (Hall, 2007, p. 349). On the contrary, some have argued that in the case of an artificial general intelligence (AGI) capable of moral reasoning similar to a human, the AI can be considered a

full moral agent and can be held responsible for its actions (Gordon, 2020). Still, the standard assumption remains the same: Agency is associated with the capacity of being held responsible for an action – an assumption that has consequently resulted in the majority of scholars resisting the idea of agency in machines or technology (Bryson, 2018; Goertzel, 2002; Johnson, 2006).

As patiency is the direct counterpart to agency, it is naturally ‘coined by analogy with the term moral agent’ (Hajdin 1994, p. 180). Just as conventional accounts understand agency as synonymous with having responsibility, the question of machine patiency so often proceeds by asking whether we have a responsibility towards certain entities and must consider them. Bryson (2018), for example, defines a moral patient as ‘something a society deems itself responsible for preserving the wellbeing of’ (p. 16).

Commonly, scholars attend to the question of moral patiency in artificial entities by asking whether we have a moral responsibility or duty towards technology or whether AI can have a claim to certain rights and respect. Gunkel’s (2012) analysis of the patiency question is once again a clear example of this underlying supposition in the technology discourse: ‘The question of moral patiency is whether and to what extent robots, machines, nonhuman animals, extra-terrestrials, and so on might constitute an other to which or to whom one would have appropriate moral duties and responsibilities.’ (Gunkel, 2012, p. 93)

Recent accounts in robo-ethics hold this same understanding: Moral patients are understood as vulnerable entities that can be ‘morally wronged’ (Gordon, 2020) and that ‘we owe direct moral duties to’ (Mosakas, 2020). A similar conception is held by Müller (2021) in his article on robot rights: ‘Full ethical agents have rights and responsibilities, while ethical patients only have rights.’ (p. 1) Or, as Howard Nye said: In order for an AI to be a moral patient, it would require us to owe moral duties such as non-maleficence to it (Nye & Yoldas, 2020).

Defining moral patiency this way consequently ties it to a question about its practical implications and consequences for our behaviour. It appears that the aim of questioning moral status across the debate is primarily to discover our duties and obligations, thereby making the moral status question ‘a means of specifying those entities towards which we believe ourselves to have moral obligations, as well as something of what we take those obligations to be’ (Warren, 1997, p. 9).

As a result, moral patiency in machines has remained mostly refused, as the idea of obligations towards robots or computers appears unthinkable to most: ‘We have never considered ourselves to have moral duties towards our machines’ (Hall, 2007). This has led to

some scholars like Joanna Bryson (2010) or Calverley (2011) strongly rejecting any existence of any duties towards machines: ‘How can pieces of metal, silicon, and plastic have any attributes that would allow society to assign it any rights at all?’ (Calverley, 2011) Others reject the idea of moral status in AI because they assume that it implies robots are given an equal legal status or even ‘given priority over the rights of any human’ (Cappuccio et al., 2020, p. 12).

But why is the question of moral patiency so often assumed to be one about how we ought to treat machines, what legal status they should hold, or what duties we have towards them? Floridi and Sanders (2004) reject this exact supposition inherent in moral status discussions as holding ‘an unacceptable assumption’ and ‘a juridical fallacy’ that equates identification (X is a moral agent or patient) with evaluation (X has responsibility or X has rights). Although Floridi (2007) has directed his criticism mainly towards the question of moral agency, I propose to re-frame the question of patiency in a similar way to allow for a different perspective.

Rather than defining moral patients as entities to whom we owe moral duties or benevolence to, I will draw from Warren’s (1997) definition of moral patients as entities that are morally considerable. How much and in what ways they should be considered would be a question of their *moral significance* and is not the objective of my questioning.

This distinction between moral status and moral significance has been similarly put forward by Goodpaster (1978) and Gruen (2003). In Goodpaster’s essay (1978), the question of moral significance is understood as a question about moral weight and priority. It is one that is relevant for practical implications such as granting rights to an entity or weighing moral priority in situations of conflicting interests. And yet, it is often applied synonymously with moral consideration by scholars. As Gruen puts it: ‘Being morally considerable is like showing up on a moral radar screen — how strong the signal is or where it is located on the screen are separate questions.’ (Gruen 2003).

To emphasise this distinction, I will define being morally considerable or giving moral consideration as suggested in an account by Thomas Birch (1993):

To give moral consideration to X is to consider X (to attend to, to look at, to think about, where appropriate to sympathise or empathise with X, etc.) with the goal of discovering what, if any, direct ethical obligations one has to X.

The advantage of this definition is that it does not make any definite claims or suggestions of what giving moral consideration should look like. After all, it is not my aim to discover *how much* we should consider technological entities but *why* we should consider them. *What* about

them makes them candidates for moral patiency and morally considerable? Is it their characteristics and ontological properties? Or is it, as the relational turn suggests, their relationships to other moral entities that matter?

Direct and Indirect Moral Status

An additional important distinction that needs to be expanded on for a thorough analysis is whether we question entities in their *direct* or *indirect* moral status. To illustrate why this distinction matters, I will introduce alternative perspectives in which moral status is used synonymously with terms like moral importance (Johnson, 2006) or moral relevance (Warnock, 1971).

In the account of scholar Deborah G. Johnson (2006), the question of moral status is formulated as a question about the *moral character* or *moral importance* of entities. Rather than only thinking of moral status in the traditional form of moral agency, Johnson encourages a view of moral status that is equivalent to the possession of moral importance. The advantage of framing moral status this way is that it becomes more easily applicable to a wider range of entities. The formulation of Johnson's (2006) argument aims to avoid any attribution of moral agency to technological entities while still acknowledging them as part of the moral realm. This is expressed in the following:

To deny that computer systems are moral agents is not the same as denying that computers have moral importance or moral character; and to claim that computer systems are moral is not necessarily the same as claiming that they are moral agents. (Johnson, 2006, p. 195)

By extending moral status questions this way, terms such as “being moral”, “having moral importance”, or “having moral status” become interchangeable with one another. But does this make for any improvement to the discussion?ⁱⁱⁱ

In a similar manner, G.J. Warnock poses the question of moral status as a question of *moral relevance*: ‘What, we may ask here, is the condition of moral relevance?’ (Warnock, 1971, p. 148).

It should be clear from earlier examples, such as the Google Search Algorithm, that technological entities significantly impact the morality of our actions. They are undeniably morally important in the sense that they are entangled with societal, political or economic issues and pose a number of ethical challenges in themselves.

Simply stating that technologies “matter morally” or have “moral relevance” is a rather uncontroversial claim. What is more controversial is the claim that they matter morally

for their own sake or that they have moral status *in themselves*. The interesting part of the discussion is not that technologies matter but *why* they do. What is the source of an entity's moral importance? This is again a question of how moral status is grounded.

The most common understanding amongst scholars is that entities have moral status when they themselves can be morally wronged or benefitted (DeGrazia, 2016) or simply when they matter for their own sake (Kamm, 2008). Especially the concept of being morally valuable for one's 'own sake' and having 'intrinsic value' is frequently underlying ideas of moral status (O'Neil, 1997). To illustrate the intuition behind these expressions, we can consider the following example: If someone destroys another person's property, for example, by setting their car on fire, their action would be intuitively considered as morally wrong. How exactly we define which actions are morally "wrong" or "right" depends on the kind of ethical framework applied and is not specifically of relevance in my account.

A utilitarian account would judge good or evil based on diminished or increased utility, a virtue ethicist would judge an action by asking whether it cultivates virtues or vices, and so forth. If we apply a utilitarian view, for example, setting the car on fire is morally wrong because it likely goes against the interest of the car's owner and diminishes its utility.

In this instance, one could say that an object like a car matters morally and has some moral importance or that our actions towards it have moral relevance. Yet the reason for its moral importance only arises because it matters to another (human) entity. We intuitively consider the action to be wrong because a human agent is harmed - not because the material of the car is damaged. It would appear counterintuitive to claim that the car *itself* is being morally wronged. The entity being harmed clearly appears to be the human owner.

While there might be good reasons to consider and protect certain objects, the moral obligation we have is not to those objects themselves. Instead, the moral importance of these entities would only arise *indirectly*. Objects like a car, necklace or even cell phone are usually only considered in moral situations because of the value and utility they provide to their human owner – not for their own sake. The point of expressions like "morally counting for its own sake" reflect the notion that an entity should have moral status beyond its mere instrumental value or utility to its user.

In the debate on technology ethics, we can find various accounts that advocate for a moral consideration of technological entities but only do so indirectly. In a paper by Cappuccio et al. (2020), for example, he argues that artificial entities should be given moral consideration as this will cultivate positive virtues, which will then result in more virtuous behaviour towards other humans. In this case, moral consideration is given for the sake of

society and other humans – not for the sake of the artificial entity. Similar arguments are put forward by scholars who advocate for the ascription of rights to robots but only do so based on a concern for future human societies and flourishing (Darling, 2016; Davies, 2011; Levy, 2009).

One of the most famous examples of an indirect argumentation for moral status is the Kantian dog: In Immanuel Kant's lecture on *Duties to animals and spirits* (1963), the act of kicking a dog is only considered immoral because it might contribute to desensitisation of cruelty towards humans. The moral status of the dog is grounded only indirectly (namely in the moral significance of the human agent) rather than in the entity itself. In Regan's (1983) account, our moral obligation to refrain from hitting the dog is referred to as an 'indirect duty', as opposed to direct duties.

In some accounts, like Johnsons (2006), having moral importance or being given moral consideration is enough to speak of the entity as having moral status. The notion of moral status in this case, however, is entirely different from the intuitive understanding introduced earlier. It is incompatible with a definition of moral status as something for the entities 'own sake' (Jaworska & Tannenbaum, 2013).

To distinguish between these two notions, I will specify between understandings of *direct moral status* or 'direct views of moral status' (Coeckelbergh, 2021) and *indirect moral status*. The main concern of my inquiry is direct moral status, as it represents the most intuitive and conventional understanding of moral status. It also reflects the notion of moral status in Gunkel's and Coeckelbergh's accounts, who similarly point out that the moral status of an entity should be at 'least partly related to something about that entity' (Coeckelbergh, 2012, p. 32).

The exact definition of an indirect or direct argument is, again, widely subjective and varies among scholars. However, there are certain terms which appear especially frequently in combination with direct moral status. One example of such terms is the concept of intrinsic value or, alternatively, inherent value (O'Neil, 1997).

Yet, defining direct moral status as the possession of intrinsic value is problematic in its own way and is bound to complicate the investigation even further. For one, the term "intrinsic value" itself opens up a whole new line of discussion about what it means to be intrinsically valuable and requires further clarification in regard to related concepts such as 'final value' (Korsgaard & Christine, 1983). Secondly, using the term intrinsic value is likely to lead down the path of a property-based argumentation. How can we determine what exactly is intrinsically valuable about an entity? When trying to define what makes an entity

intrinsically valuable, one easily becomes tempted to search for certain characteristics to justify said value. Therefore, I am hesitant to apply a definition of direct moral status based on concepts such as inherent or intrinsic value. As scholars such as Rick O’Neil (1997) have suggested, I find the notions of moral status and intrinsic value to be best understood separately from one another so that neither is a condition for the other.^{iv}

Describing moral status as *direct* or rooted within the entity itself should not necessarily require it to be rooted in the entity’s properties or ontology. Otherwise, the only possibility for direct moral status ascription would be a property-based argument, and any investigation into alternative argumentations would be redundant.

Coeckelbergh (2012), who provides a relational view of moral status for example, claims to provide a direct justification of moral status and even criticises other relational arguments made by Cappuccio et al. (2020) or virtue ethicists for their indirect ascription of moral status (Coeckelbergh, 2012). He argues *against* an indirect argumentation for moral status and claims that an ‘indirect argument does not account for our intuition that moral status should be about the entities in question and not about us’ (Coeckelbergh, 2012, p. 6).^v In doing so, he challenges one of the common suppositions that relational arguments only attribute moral status in an indirect way – a claim that will later be investigated further in a review of relational accounts.

For the purposes of my inquiry, direct moral status will be understood in the context of Gunkel’s (2010) and Coeckelbergh’s (2012) definitions. Direct moral status is thereby defined as a justification of moral status that is at least partly rooted in the entity. The advantage of conceptualising moral status this way is that it does not require us to recognise technological entities as moral patients *in themselves*. As the relational analysis will later make evident, an ascription of status based on concepts like ‘intrinsic value’ or ‘having one’s own sake’ becomes meaningless anyway when the status of an intra-relation is thought of as constituted by each other (Verbeek, 2011).

Property-Based Theories

Having defined both the scope of technologies and my understanding of the terminology around moral status so far, I will now begin the second part of my inquiry. As outlined earlier, this section will be concerned with how moral status is grounded by comparing two contrasting argumentative strategies: Property-based and relational justifications of moral status, specifically the social-relational argument made by Coeckelbergh (2012). This will further allow to illustrate a broader contrast between the

perspectives of individualist accounts of morality within Western modernity and Eastern views of moral status (Ess, 2015).

The most commonly used approach towards the moral status of technologies to date is an inquiry based on their possession of a certain property. It is also how we generally structure our moral understanding in practical interactions: If we come across an entity that is found to be lifeless and incapable of experience, such as a rock on an open field, then we would find it morally permissible to kick the rock into the distance without worrying about harming it. If we encounter our neighbour's dog or children, on the other hand, we know to refrain from kicking or harming them because we believe them to possess properties like sentience and a capacity for pain. Entities are categorised by us on a daily basis according to their properties – whether we are aware of it or not – which then pre-structures the kind of behaviour we find morally permissible or not. This kind of approach is known as the 'properties approach' (Gunkel, 2018; Tavani, 2018), or alternatively as the 'mind-morality approach' (Gerdes, 2016), 'standard approach' (Coeckelbergh, 2012) or the 'organic approach' (Torrance, 2008).

From the perspective of a property-based account, moral status is something innate and certain to the entity itself. Property-based arguments are often found to be *direct* because moral status is grounded intrinsically within the ontological conditions of said entity (Søraker, 2007).

Within the field of technology ethics, scholars predominately question properties such as the possession of sentience or mental states (Coeckelbergh, 2012). While there remain disagreements among scholars about which property is the relevant one, the underlying consensus is the same: If an entity possesses x characteristic or property or is likely to do so in the future, then the attribution of moral status is justified (Müller, 2021).

The formal structure of a property-based argument has been laid out by Coeckelbergh (2012, p. 14) as the following:

- (P1) Having property p is sufficient for moral status s
- (P2) All entities of class c have property p
- (C1) Therefore, all entities of class c have moral status s
- (P3) Entity e is of class c
- (C2) Entity e has moral status s

To successfully follow this argumentative structure, two steps are needed: First, it has to be established that the entity even possesses the relevant property in the first place, and secondly, one has to demonstrate that the property is morally significant (Samuelsson, 2008).

Especially the justification of why a certain characteristic is more morally significant than the other, will later turn out to be one of the main difficulties within the approach.

It is important to note that the moral status ascription is assumed to take place from the perspective of an impartial observer, isolated from any social context of the world (Coeckelbergh, 2012, p. 17). The aim of the property-based approach is to determine properties that are intrinsic to thereby gain insight into the true nature and ontological reality of an entity (Coeckelbergh, 2012, p. 5). Therefore, the properties in question are not understood as mere appearances but in the Kantian (1788) sense of ‘the thing itself’. For questions of moral status, this means that it does not matter whether an entity appears to have the property of sentience based on its outer appearances or behaviour: Only objective knowledge and evidence can determine ‘the thing in itself’ and, consequently, moral status.

A similar assumption is held by proponents of moral individualism, who argue ‘that only intrinsic properties can be status-conferring and give rise to agent-neutral moral reasons’ (McMahan, 2005, p. 357). It further resembles a typical viewpoint for Realism, which holds that entities have an objective reality independent of our perceptions or beliefs (Tollon & Naidoo, 2021). Objectivists favour an unambiguous answer about what is morally right, wrong, good and bad. Hence, some scholars refer to property-based views as *realist* or *objectivist* (Coeckelbergh, 2012, p. 80).

In the example of robots, a mere display of intelligent behaviour, problem-solving skills or conversational competencies would not be sufficient to determine whether it is actually intelligent. It might only be giving the illusion that it is, without having any of the real sense of understanding required for true intelligence. Objective reality (what things really are) is hereby separated from perception (what things appear to be).

In this way, the property-based view makes a strict distinction between appearances and truths, between objective reality and perception. This is characteristic of the metaphysical views within Western philosophies, which emphasise values of rationality and objectivity. The emphasis on these values can be largely blamed on the historical context in which the property-based approach gained traction in. Within the Enlightenment movement of the 18th century, philosophical inquiries became deeply influenced by the scientific progress and breakthroughs happening throughout (Coeckelbergh, 2012, p. 166). In a century of philosophy par excellence, science provided philosophers with new tools to empirically observe, measure and analyse an entity’s characteristics. As a result, moral status ascriptions took on a scientific character and proceeded by trying to find rational justifications for moral status. Similarly, Coeckelbergh describes the property-based approach as a scientific rather

than philosophical approach and characterises it as an ‘outgrowth of the habit of rational thinking’ (Coeckelbergh, 2012, p. 166).

The Western rationalist tradition undeniably still shapes the majority of moral status ascriptions within technology ethics today: The most popular criteria to ground the moral status of technology in is sentience or consciousness (Himma, 2003; Johnson & Verdicchio, 2018; Mackenzie, 2014; Mosakas, 2020; Tomasik, 2014; Torrance, 2008). Alternatively, some scholars within the field of machine ethics have proposed characteristics like rationality (Laukyte, 2017), interests (Basl, 2014), autonomy (Calverley, 2011) or moral reasoning (Malle, 2016).

For now, it remains disputed whether sentience, rationality or interests are present in today’s technologies or not. Yet, under the property-based view, it would hold that if a certain property can be objectively detected and verified, then moral consideration is appropriate: ‘As soon as AI begins to possess consciousness, desires and projects, then it seems as though they deserve some sort of moral standing.’ (Sparrow, 2004, p. 203)

From a property-based perspective, a determination of moral status seems straightforward at first glance. The property-based argument would examine whether an entity meets a certain criterion (such as a capacity for suffering, interests or experience), which then concludes whether the entity qualifies for moral patiency or not. If sentience is chosen as the foundational requirement for moral status and is defined as the capacity for phenomenal experiences and feelings (DeGrazia, 2016), then most AI systems would be excluded from moral patiency by authors like Giroux and Larue (2015), Joshua Shepherd (2018) and Johnson and Verdicchio (2018).

A similar stance would be taken by scholars that find cognitive abilities to be crucial, as algorithms and AI for the foreseeable future do not have ‘any of the objective cognitive prerequisites that could possibly identify them as persons or moral patients’, such as self-awareness, autonomous decision, motivations, or preferences (Cappuccio et al., 2020, p. 3).

One of the few criteria that would allow us to define search algorithms and other intelligent technologies as moral patients was proposed in Floridi’s *Information Ethics* (IE) (Floridi, 2008). Here, the property that is found to be more fundamental than suffering or sentience is entropy which applies to all informational objects (Floridi, 2008, p. 11). In this case, a moral action can be anything that contributes to the destruction of an informational being and hereby causes a higher level of entropy. As a consequence, an algorithm can be a moral patient as an action like a malware attack would cause higher levels of entropy (Floridi,

2008). However, choosing a universal, all-encompassing criteria for moral status like this comes with its own set of problems, as a review of a critique by Gunkel (2012) will show.

Problems With a Property-Based Approach

Oppositions against the property-based approach have been voiced by a number of scholars (Coeckelbergh, 2014; Danaher, 2020; Gunkel, 2012; Tavani, 2018). An example of such opposition is found in the account of Thomas Birch (1993), who argues that the employment of any property is ‘an act of power over (...) those others who turn out to fail the test of the criterion’ (Birch, 1993, p. 317).

One of the most extensive critiques of the property-based approach, however, can be found in David J. Gunkel’s *Thinking Otherwise* (2007) and *The Machine Question* (2012). His accounts will, therefore, be used as my primary point of reference to identify the initial weaknesses of the property-based argument.

Gunkel's criticism broadly comprises four areas of concern: The first is a concern with the underlying structure of the method itself, which seemingly always proceeds through so-called ‘mechanisms of inclusion and exclusion’, thereby resulting in the exclusion of certain entities from the moral sphere (Gunkel, 2012, p. 160). The remaining points of weakness Gunkel identifies are the epistemological challenges arising from property-based arguments (‘problems of detection’) and the difficulties of defining and deciding on certain properties. He refers to the latter as ‘problems of determination’ and ‘problems of definition’ (Gunkel, 2012).

What exactly Gunkel means when he speaks of ‘mechanisms of inclusion and exclusion’ can best be illustrated in a review of the development of the moral circle over time. Throughout most of history and Western philosophy, moral status ascription took place in an anthropocentric tradition. In the ordinary use of the term, *anthropocentrism* refers to views in which only (yet not necessarily all) humans have moral status (Samuelsson, 2008). Such views are found throughout the works of well-known philosophers like Aquinas, Descartes and Kant. In their accounts, human beings are the only entities that are considered to have inherent value and are consequently the only entities with moral status (Steiner, 2010). The intrinsic value of human beings was justified by these philosophers through properties such as rational thought and reason (Kant, 1785). It resulted in the exclusion of any non-human animals that were found to have ‘no reason at all’ (Descartes, 1637).

The discourse on animal rights was one of the first to significantly challenge the categorical exclusion of non-humans: In his book *Animal Liberation*, Peter Singer follows Jeremy Bentham (1789) to ask whether an animal has ‘the capacity to suffer’, and thereby interests and preferences which warrant its moral status (Singer, 1975, p. 8). Others, like Tom Regan (1983) argued for the moral consideration of animals from a deontological standpoint, pointing out that they are ‘subjects-of-a-life’ the same way humans are.

According to environmentalist accounts, this expansion of moral status to non-human animals still did not go far enough: While animals were now part of the moral circle, other species, such as plants or microorganisms, remained excluded in the accounts of Regan (1983) or Singer (1975). As an alternative, accounts such as Leopold’s ‘land ethic’ proposed a biocentric perspective to ‘enlarge the boundaries of the community to include soils, waters, plants, and animals, or collectively, the land’ (Leopold, 1966, p. 39).

Taylor’s (1986) argument in *Respect for Nature* is another such attempt to find a more inclusive criterion. It employs properties such as ‘having a good of its own’ and ‘teleological functions’ to include all natural entities as moral patients (Taylor, 1986, p. 124). While such eco-centric arguments seem to be more inclusive and apply to all members of the natural environment, they still take an exclusive approach, according to Gunkel (2012).

In Taylor’s account, for example, only natural entities such as animals or plants are eligible for moral status. Meanwhile, any non-living, non-natural entities like machines or artefacts are excluded since they do not have a teleological function of their own but only derivative purposes given to them by humans: ‘The ends and purposes of machines are built into them by their human creators’ (Taylor, 1986, p. 124).

The most universal and all-inclusive accounts attempt to overcome even these last barriers of exclusion and attempt to extend moral patiency to all entities, including artefacts and synthetic organisms (Basl, 2014; Floridi, 1999). Onto-centric arguments made by Basl and Sandler (2013) suggest that even artefacts and machines should be morally considered as they are teleologically organised towards a certain end or purpose. Whether those purposes are only derivative of human interests does not undermine the fact that machines or artefacts have ‘teleo-interests’ that can be harmed or promoted (Basl & Sandler, 2013).

Another onto-centric view was proposed in Floridi’s (1999) IE, in which consideration should be extended to anything that exists as a body of information. The range of moral patients includes all entities ‘from paintings and books to stars and stones, anything that may or will exist’ (Floridi, 1999, p. 43). Whether the onto-centric arguments go too far and diminish any morally relevant differences between entities can be reasonably questioned.

Yet considering that once even slaves were considered non-human and were excluded from moral status, it seems that the overall effort for more inclusion is an amicable progress.

Gunkel (2012), however, criticises both inclusive and exclusive arguments since they are structurally and systematically the same in their approach: They all ultimately proceed by finding a criterion that constitutes the boundaries of moral status. According to Gunkel (2012), the problem is not the debate about who should or shouldn't be included in the moral circle but the strategy itself. Whenever moral status is granted based on properties, some entities will always remain excluded (Gunkel, 2012).

One could argue that all-inclusive accounts like the ones of Floridi (1999) or Basl (2014) don't make any exclusions – any entity would qualify for moral consideration in IE simply by being a body of information. However, Gunkel (2012) considers this to be equally problematic as it reduces all entities to their shared common property and to a level of abstraction: 'Exclusion is a problem because it calls attention to and fixates on what is different despite what might be similar. Inclusion is a problem because it emphasises similarities at the expense of differences' (Gunkel, 2012, p. 162). To underline this, Gunkel refers to Birch (1993), who characterises both inclusive and exclusive criteria as an act of violence and power against others (Birch, 1993, p. 317).^{vi} Despite being more inclusive, ontocentric arguments ultimately fall back on the same established structures and protocols within moral philosophy and remain centrist (Gunkel, 2012, p. 152). The only difference is that in the case of Floridi's IE, it is a more inclusive ontocentrism rather than anthropocentrism or biocentrism employed. A similar criticism is expressed in Introna's (2007) ethics of hybrids, in which he urges us to 'admit that any attempt at humanistic moral ordering is it egocentric, anthropocentric, biocentric (...) or even ecocentric will fail' (Introna, 2007).

The second criticism voiced by Gunkel (2012) is founded on the epistemic uncertainty we have about the existence of any property. It is similarly reflected within lines of moral irrealist thought (Hare, 1996; Harman, 1986; Stevenson, 1994). As already discussed, a property-based account establishes a distinction between appearance and truth: It creates a gap between what properties we perceive an entity to have compared to the objective truth about its properties, which science reveals (Gunkel, 2012, p. 139). Advanced robots may very well exhibit mental states and intelligent behaviour, yet the common view amongst computer scientists and various scholars is that mere appearances do not reflect the actual possession of these properties (O'Regan, 2007; Searle, 1980; Torrance, 2008). Whether a robot is capable of showing 'outward behavioural signs of distress' (Torrance, 2008, p. 499) does not indicate an equivalent inner sentient state.

Similarly, Kevin O'Regan (2007) argues that we may design robots capable of screaming and acting in pain, yet the machine may never actually know “what it is like” to have pain (O'Regan, 2007). Truly having a mental state requires the experience of what something is like or so-called phenomenal consciousness (Block, 1995).

The underlying idea of an appearance/reality distinction goes back to its earliest articulations in Plato's ‘Allegory of the Cave’ (Reeve, 2004) and is reflected in later European philosophies of thinkers like Descartes (1637) or Kant (1788). In the case of Plato's cave, what presents itself as a shadowy figure on the wall is merely a deceptive image. It differs from the figure's real nature as it would be perceived in direct sunlight outside the cave. To recognise this ontological difference, one would need to go beyond the mere appearances of a thing (in this case, outside of the cave) and instead have access to the reality of the thing in itself (Gunkel, 2012, p. 139). For technological entities, this means that properties can only be attributed if they represent an objective truth about the entity.

But for criteria such as consciousness or sentience, this objective truth is just as unattainable in modern science today as it was hundreds of years ago. Psychologist Stuart Sutherland already declared thirty years ago that consciousness is unlikely to ever be fully understood: ‘It is impossible to specify what it (consciousness) is, what it does or why it evolved’ (Sutherland, 1996, p. 95). Computer scientist Ray Kurzweil voiced a similar concern in his book *The singularity is near*: ‘Consciousness (...) is not scientifically reliable’ (Kurzweil, 2005). Any evidence for material and scientific truths about sentience or consciousness always remains restricted to our human mental reality as we perceive it.

Consciousness is already hardly understood in itself and even less so objectively verifiable in other entities. This cartesian doubt manifests itself in different forms of scepticism and gives rise to what Thomas Nagel (1974) coined as ‘The other mind problem’ – How can we ever determine whether other beings have consciousness, thoughts, sentience or autonomy when we only have access to our own mind?

We might even doubt if other humans have a mind – what if they are ‘philosophical zombies’? – a term suggested by Chalmers (1997) to describe humans that are indistinguishable from us but lack mental states. The epistemological challenge seems unavoidable whenever we use intrinsic properties to establish moral status: ‘If animals (or machines) have an inner mental life, how would we ever know it?’ (Gunkel, 2012, p. 117).

In addition to the epistemic uncertainty, there persists a problem in finding a univocal definition for properties like consciousness, autonomy or sentience among scholars. One of the most ambiguous and still widely applied terms is probably consciousness – which

scientists barely have an understanding of – let alone a universal definition of. As Velmans (2000) points out: ‘(Consciousness) means many different things to many different people, and no universally agreed core meaning exists’ (Velmans, 2000, p. 5).

Lastly, the properties approach faces the problem of determination (Gunkel, 2012): A fundamental difficulty is deciding which property should determine moral patiency in the first place. What matters for the moral status of an entity – intelligence or sentience? Should we apply Regan’s (1983) criterion of being the ‘subject-of-a-life’ or is having an interest (Singer, 1975) the more suitable property? What makes one criterion better than the other? Even if we find a way of proving that machines and other technological entities are intelligent or sentient, the question remains as to why it should justify their moral status. Scholars continue to disagree about which property is actually ethically relevant for moral status. And as the century-old discourse on moral status suggests, the lack of agreement on these questions is unlikely to resolve itself anytime soon.

Due to several philosophical difficulties (problems of determination, detection and definition) and a problematic centrist structure, Gunkel (2012) ultimately concludes that the property-based view is incapable of answering the question of moral status for machines. As an ethical theory, Gunkel (2012, p. 1) argues that the standard view is not agile enough to respond to the unique challenges modern technology poses.

However, I suggest that these are not the only grounds on which the property-based account should be dismissed. It is not necessarily the lack of clear definitions and epistemic uncertainties that make property-based accounts insufficient in dealing with modern technologies – rather, it is the kind of worldview sustained by it. The property-based approach is unfit to question the moral status of today’s technology because it rests on a set of modernist assumptions that can no longer be upheld. Not only is it hard to prove the existence of properties or find univocal definitions for them, but moreover, it is the implications of the property-based perspective that turn out to be problematic. Resulting from an objectivist view of morality is an atomistic representation of the world and the entities in it, which I find to be untenable in today’s environment.

To demonstrate this point, I will recall some of the underlying assumptions of a property-based argument in the next section. From this emerge a number of problematic implications for our view on selfhood, technological entities and moral status ascriptions overall which provides further reason to move on to a different framework.

An All Too Modern Problem

Different philosophical theories generally come with certain kinds of metaphysical commitments and assumptions – and it is no different in the case of various moral frameworks. Previously, I have already touched on the parallels between the perspectives of property-based and relational accounts to the contrasting perspectives of Western and Eastern philosophies.

The emphasis on values like objectivity in the property-based approach is just one example in which these similarities are visible. Much of the assumptions underlying property-based accounts are representative of modernist values that are typical within classical Western philosophy (Ess, 2015). The very search for an objective characteristic for moral status – the very idea that moral status is something that can be objectively measured – presupposes a separation between reality and appearances that has been central to Western philosophy since Plato’s allegory of the cave (Reeve, 2004). It manifests itself in the well-known and never-ending pursuit of ontological questions about the true nature of entities: ‘What is the being behind the beings?’ (Ames, 1998, p. 2).

The possibility of even posing questions of this kind presupposes a dualistic worldview or, as Søraker (2007) describes it: *A two-world metaphysical view*. The dualistic worldview of Western philosophy impacts both how we understand selfhood as well as how we view entities in the world around us (Søraker, 2007).

In the following, I will expand on both of these understandings – namely that of selfhood and reality – and question them in the context of modern technologies. Doing so further reveals how significant the context in which moral actions are understood truly is. Throughout my reflection on the modernist understanding, the additional shortcomings and weaknesses of a property-based view become increasingly obvious.

The first prominent feature of the modernist view is its individualistic understanding of selfhood. In order to achieve a successful and correct identification of an entity’s properties, the entity would need to be analysed from an entirely objective point of view. The property-based view assumes that moral status ascriptions take place from the perspective of a rational, disembodied human who can analyse and dissect the reality around him. The understanding of self that is hereby established reflects an outdated conception of a rational, autonomous self that is characteristic all throughout modern Western thought: Within continental philosophy, the idea of human beings as individual unique selves precedes any deliberations of morality (Wassmann et al., 2013, p. 233). It is an understanding of selfhood

that, as Thomas Nagel might say, takes ‘a view from nowhere’, detached from the actual, material world (Nagel, 1989). Consequently, the moral agent is taken outside of the context in which a moral action occurs in rather than being placed in it. Understanding moral status ascriptions this way is problematic for numerous reasons:

For one, applying the assumption of an objective moral agent is burdensome in that it is clearly detached from the reality of moral status ascriptions. Our judgment of which entities possess moral status has never taken place as a formal procedure in which we first analyse the properties of an entity and then ascribe moral status afterwards. The other minds problem is a concern that is mostly relevant in theory— in practical circumstances, however, our epistemic uncertainty about the existence of consciousness in other humans has not stopped philosophers from widely assuming it. In a majority of accounts, humans are automatically granted moral agency based on their possession of sentience, despite the logical possibility that our fellow humans are merely ‘philosophical zombies’ – an argument that was famously discussed by Chalmers (2014). Similarly, our lack of knowledge about technological entities and their properties has not hindered us from ascribing consciousness or mental attributes to robots and machines. Countless studies on human-robot interactions have demonstrated that knowing about a robot’s lack of sentience does not affect participants’ abilities to sympathise or even ascribe mental states to them (Kwok et al., 2018).

The problem I intend to point out here is not necessarily that the modernist view of self is unrealistic. After all, solely rejecting an ethical theory because it is detached from reality does not make for very strong grounds for dismissal. The bigger problem is rather the view of technology resulting out of it. Analysing the moral status of intelligent technologies from the place of a disembodied, detached agent is unsuitable in that it risks overlooking a crucial aspect of modern technology: Its interactivity and interwovenness with selfhood itself.

Technologies and their characteristics can hardly be understood if they are approached as passive, lifeless objects. A property-based approach establishes a view of technology as entities that exist in separation from us: They become mere objects of scientific inquiry, ready to be analysed in their features. Thought of this way, the property-based perspective offers an approach to moral status that appears more like a scientific method rather than a mode of philosophical inquiry (Coeckelbergh, 2012). Attempting to analyse the ontology of technological entities this way does not only seem unsuitable but entirely unfeasible.

As laid out in numerous examples earlier, technologies like smartphones or algorithms not only change our perception of the world around us but also shape who we are and how we identify. They have become an integral component of society overall and

inseparable from individual selfhood itself. The Western notion of an indivisible individual as the designated identity of human beings has been increasingly challenged since the emergence of digital forms of selfhood. Personal identity is no longer confined to our physical bodies but extends beyond the material world into the virtual sphere.

The phenomena of selfhood going beyond our bodily boundaries has prompted numerous new conceptions of self and identity within philosophical literature: Authors like Biocca and Delaney (1995, p. 58) have discussed the transformation of selfhood through technologies like VR as ‘a medium for the extension of body and mind’, and Clark and Chalmers (1998) have extensively challenged the distinction between mind, body and environment in their *Extended mind theory* (1998). Others have articulated the notion of a multi-faceted self through terms like ‘telepresence’ (Minsky, 1980) or ‘remote presence’, which can occur in both subjective and objective ways (Draper, 1998).

The most straightforward examples of virtual identities today are avatars in video games or social media profiles. In these cases, the username or avatar is perceived as a representation of our own self and is often a direct component of personal identity. Research indicates that insults to a virtual character are perceived as no less harmful than ones to the physical self by the victim (Wolfendale et al., 2007). The morality of such actions is just as significant in the virtual as in the physical sphere, yet it is often perceived as less harmful by those committing them.

With recent developments such as Apple’s launch of “Vision Pro glasses” and other AR technologies, virtual presence is no longer just a digital representation of self but merges with our physical presence. An avatar in a video game can somewhat be thought of as separated from us and as confined in the virtual world – it is confined behind the screen of the computer. Yet, with AR technologies or implanted chips, the borders between the virtual and material world slowly but surely disintegrate. With AR glasses, it is no longer necessary to press a button or type on a keyboard to communicate our needs to the machine – simply moving our eyes is enough (Asaad, 2021).

While all technological developments point to a further dissolution of the boundaries of the individual, the property-based view holds onto the perspective of Western metaphysics of entities as stand-alone things. An approach of this kind ultimately fails to acknowledge the concrete realities of today’s interdependencies and instead establishes an atomistic representation of the world (Callicott, 1989).

While relations and interactions between entities can still occur in this view, the individuality and organic unity of them is always strictly preserved. The relation between the

observer and the observed entity can only be understood to exist on a surface level, thereby protecting the unity and identity of each entity from dissolution: When the relationship between them is dissolved, the previously related entities remain intact (Rosemont & Ames, 1998, p. 24).

The Downfalls of an Atomistic Worldview

Not only does this constitute a view of individual entities as separate from another, but it also fosters a practice of allocating entities into separate categories. Identifying entities based on a pre-determined property theoretically allows us to divide the world into neat categories: Those who have intrinsic interests (or a capacity to suffer) and are thereby part of the natural sphere, those with derivative interests as members of the technological sphere, and so forth.

The practice of categorising entities into distinct and contrasting classes is one with a long history. An especially popular example is the *nature-culture* dualism, which was prominently held by philosophers during 17th-century Modernity: The nature/culture dichotomy describes a worldview in which the natural and the social were separated into two ontological categories (Collingwood, 1960).^{vii} What exactly constitutes the distinction between the two has been defined in various ways throughout history. Yet, a common assumption is that culture is that which is man-made or constructed (e.g., in the form of artefacts), whereas nature is equated with the external, self-developing environment (Haila, 2000).

By now, the dualistic conception of nature vs. culture has been problematised in numerous accounts and is widely considered to be flawed (e.g., Heidegger, 1996; Ingold, 1996). Ingold, for example, argues that there can be no distinction made between what is cultural/social and what is natural: ‘We can no longer think of humans as inhabiting a social world of their own, over and above the world of nature in which the lives of all other living things are contained’ (Ingold, 1996, p.87). Instead, humans, animals, and the plants on which they depend for a livelihood must be regarded as fellow participants in the same world – a world that is both social and natural at the same time.

Just looking at our daily lives, it seems reasonable to argue that animals are legitimate members of our social world: They are part of the social as pets and companions, as partners in labour processes, constitute our means of consumption and are competitors for habitats (Benton, 1993, p. 18).

At the same time, humans themselves are creatures of nature and inherently part of the natural, which makes a classification of culture as ‘unnatural’ seemingly odd. What exactly determines the identification of something as truly natural, given the countless overlaps in both nature and culture?

It may appear that this is just once again the problem of definition, as Gunkel already articulated. But the problem I intend to reveal is not just a difficulty of defining what it means to be “natural” or “technological”. It is more so the kind of worldview arising from this view that I find to be deeply problematic in the modern digitalised world.

When it comes to technological artefacts and digital entities, the same kind of dualistic thinking is at play in the majority of property-based accounts. Within the moral discourse, it is rather common that technology is characterised as something detached and external to us: Algorithms and artificial entities are conceived separate from the human inhabitants of the “real” world and are instead inhabitants of the virtual sphere. The technological/natural distinction is often strengthened further by conceptualisations of technologies as lifeless instruments which do not share the features of natural and biological organisms and, thereby, have no inherent value (Regan, 1983; Singer, 1975).

An example of the real/virtual dichotomy being upheld within literature is the 1990 Manifesto *Across the Electronic Frontier* (Kapor & Barlow, 1990). By drawing on William Gibson’s (1984) novel *Neuromancer*, the authors propagate a hard distinction between the real world as ‘meat-space’ opposed to the virtual world as ‘cyberspace’ (Kapor & Barlow, 1990). In this case, cyberspace is defined as a place where all online interactions occur and is conceived to be inherently separated from the boundaries of the physical world (Kapor & Barlow, 1990).

Many philosophers of the Frankfurter Schule (e.g., Herbert Marcuse, Eric Fromm, Theodor Adorno) hold similarly dualist views. Habermas (2019), for example, suggests a separation between ‘life worlds’ and ‘systems’. Like most portrayals of the technological sphere, Habermas’s characterisation of ‘systems’ has an intentional negative undertone. The term is applied to warn about the increasing colonisation of the human lifeworld by systems of technology, which could deteriorate the character of social relations into that of ‘thing’ (Habermas, 2019, p. 186).

Another case of the technological/social dichotomy at work is Feenberg’s (2010) argument in his book *Between Reason and Experience*. Here, Feenberg states that our democracy and human-political life have intervened into the technical sphere – again presuming that they were ever separated in the first place.

Approaching moral status from a property-based perspective perpetuates precisely this separation of technological entities from other candidates of moral status. Rather than acknowledging the interrelatedness of the current world, entities are confined within their respective spheres: The social (made up of humans with linguistic and communicative abilities), the natural (consisting of all natural, living entities) and the technological/virtual (made up of artefacts and artificial entities). The moral landscape is investigated as if it were a ‘collection of objects’ and is treated as merely an aggregate of separate individual entities (Callicott, 1989).

Why is a worldview of this kind problematic in the first place? There are certainly arguments in favour of the modern, objectivist perspective: It allows entities to have universal, objective, and observer-independent moral values in themselves while avoiding the dangers of relativism. Further, some might even argue that certain distinctions and boundaries between entities are inevitable and necessary. Doing so protects the special role of humans within morality and prevents them from being equated to other non-living entities.

I do not claim that any dualistic distinctions are unnecessary and should be abandoned. Yet, the dualistic worldview underlying the property-based account simply does not make for a good starting point for moral status inquiries – especially when it comes to an inquiry into intelligent technologies. Viewing entities in compartmentalised ways is problematic because it, firstly, does not acknowledge the increasing interwovenness of the social and natural with the virtual/technological world. Secondly, it does not offer a sufficient answer in the case of hybrid entities.

My first point is that holding up a rigid separation between the technological and social world makes it virtually impossible to acknowledge the ways in which humans have become intertwined with both the virtual and the material sphere. The common dictum “all technology is social” is supported by a large and growing body of accounts which cannot be done justice here, yet I hope to point out a few of the most prominent arguments.

From the perspective of *actor-network theory*, society must be understood as a network made up of both humans and non-humans, in which functions and roles are redistributed amongst actors (Latour, 1996). An example analysed by Latour (1996) is the door-stopper, which offers a technological solution to the once-mechanical human task of closing a door. As a human obligation is delegated to the technological artefact (the door-closer), it becomes part of the social collective (Latour, 1996).

Further, biotechnologies and medical devices do not only break down the boundaries between technology and nature (e.g., in the case of genetically engineered organisms) but

have been argued to also dissolve the category of the social (Rabinow, 1992, p. 242). With the advent of ICT technologies and social media, the dissolution of the social and technological is now undeniable: New relations or political movements can arise from social platforms; texting can create new kinds of languages (Crystal, 2008) and algorithms like in the example of Google can impact entire value systems.

The human self is no longer limited to a bodily presence but is accompanied by a virtual presence in the form of social profiles, avatars, usernames or even online banking ID numbers (Verbeek, 2015, p. 217). Where the virtual world was once only accessible through an interface or monitor, we now no longer have to “step” into the cybersphere. Instead, we are fully immersed and embedded in its environment.

Secondly, we can consider the increasing cases of hybrid entities which combine both natural and technological properties. The term “hybrid” can be misleading as it has been applied in several different ways and is often associated with the image of humans as cyborgs (e.g., Clark, 2001; Haraway, 1991). For the sake of simplicity, I will refer to hybrids as any entities that possess both artificial and natural elements and/or functions.

An example is the case of artificial coral reefs, which are currently frequently used to sustain biodiversity in oceans and enable the growth of natural corals (Chen et al., 2015). Even though they are artificially created by humans, they do not simply serve human purposes as instruments but are part of the biosphere and contribute to its overall good (Holy-Luczaj & Blok, 2019). Although they technically have “derivative interests” and did not evolve naturally, they clearly belong to a natural ecosystem at the same time. They are just as much technological as they are natural. Cases like these illustrate just how difficult it is to categorise novel entities in the first place.

In the property-based approach, the world consists of boundaries and categories: It does not allow us to include artefacts and technologies into the social and human community as long as they do not pass the test of certain criteria. The previous examples, however, show that technologies are already nonetheless intertwined with the social and natural, regardless of the properties they possess.

Relational Theories

Somewhat of a solution to the modernist suppositions at work in property-based accounts is offered by a relational perspective. Although there is no clear, univocal definition of what a “relational theory” entails, a shared element across relational accounts is that they focus on the dynamics between entities rather than on the entity itself. Unlike a property-

based view, where the individual is the primary unit of analysis, the relational account prompts a shift to focus on the relationship between them. What matters for moral status is no longer an intrinsic property of an individual entity but an extrinsic or interactive element *between* two entities. As Coeckelbergh describes it, the idea of a relational account is that moral status is neither grounded in the patient nor in the agent but that it is grounded in the relation between them (Coeckelbergh, 2012).

Relational arguments hereby stand in direct contrast to a property-based, individualistic account of morality. Rather than putting the rational, objective human agent at the centre of moral status ascriptions, it allows us to approach ethics as a matter of relational selves. In recent debates on machine and robot ethics, the proposal for a more relational understanding of technologies has been described by some as a *relational turn* (Gerdes, 2016; Gordon & Gunkel, 2022).

Some well-known applications of the relational perspectives to technology recently are Coeckelbergh's (2012) social-relational arguments and Gunkel's (2007) proposals in *Thinking otherwise*, yet the scope of relational accounts is constantly expanding.

Other recent advocates for a relational perspective include Danaher (2020), who suggests so-called 'ethical behaviourism'. In this particular theory, moral status would be granted as long as the observable performance of an entity is equivalent to that of a human.^{viii}

Relational approaches to morality, however, are not at all a recent turn taken in ethical discussions: Elements of relationalism have been present in Eastern views such as Confucianism (Li, 2014) and are reflected in the communitarian thinking of Aristotle's work *Politics* (1984) and in contemporary communitarian theories (MacIntyre, 1984).

Throughout the twentieth century, relational thinking was taken up within feminist care ethics (e.g., Donovan & Adams, 1996; Gilligan, 1982), which emphasises human relationships, as well as in environmentalist accounts such as Naess 'deep ecology' (1995).

The only difference in environmentalist theories is that it is not human interpersonal relations that are emphasised but natural relations between the members of an ecosystem (Naess, 1995). Furthermore, we find relational thinking reflected all across the school of philosophical phenomenology. Phenomenological accounts embody some of the core elements of relational thinking by placing an emphasis on human embodiment and interdependency.

Considering the rich variety of relational frameworks, Coeckelbergh's (2012) and Gunkel's (2022) description of a relational turn appears to be more fittingly characterised as a

renewed focus: It is a return to a concept that has been relevant within ancient Western and Eastern views for quite some time rather than a novel strategy.

In my previous overview of indirect and direct notions of moral status, I have noted that relational arguments tend to be associated with an indirect ascription of moral status. Those who define moral status based on concepts like intrinsic value, might even find relational arguments to be entirely incapable of direct moral status attribution. This, however, would be an over-generalisation and fails to realise the various nuances amongst relational accounts.

Some accounts, which are mistakenly referred to as being relational, are certainly only ascribing indirect moral status. An example is the account of Darling (2016), who ascribes moral status to robots based on their importance for the future of human societies. The moral status ascription, in this case, is entirely rooted in the interests of the human subject and only barely touches on the relationships between human and robotic entities.

Accounts such as these have been criticised by Coeckelbergh (2012) for only being relational in a ‘shallow’ way and, hence, are not the kind of relational argument considered in my analysis.

An example of a non-shallow relational account is Naess ‘deep ecology’ (1995): Contrary to some other environmental accounts, moral status ascriptions towards plants, nature or animals are not just done for our own sake or humanity's future (Warren, 1997). Instead, Naess (1995) recognises the moral status of natural organisms beyond their instrumental value to us and understands them as a deeply valuable part of our shared natural world. In this case, the moral status of nature is neither rooted in human interests nor in the entity's individual properties. Naess (1995) does not ground moral status in either one or the other entity but rather in the complex interrelations and interdependences between them as part of a larger collective.

My following use of the term “relational argument” is primarily meant to denote relational understandings of this or a similar kind rather than those who follow a “shallow” or indirect relational argumentation.

There are several other cases in which scholars speak of relational accounts of technology, however not all of them constitute a relational account of morality. An example are actor-network theories (Latour, 1993), which have specifically set out to analyse our relations to artefacts and technologies. Others include post-phenomenological accounts, such as Idhe's (1993), that investigate the *mediating* quality of technology. In these cases,

however, the term relational is only applied for the sake of analysing the characteristics of the entity and not for an analysis of its morality.

Similarly, phenomenological accounts are considered to be of relational nature, yet they are not per se a means of moral or ethical inquiry. If anything, moral phenomenology is applied towards metaethical questions to study the what-is-it-likeness of moral experiences – without making normative judgments about whether the moral experience is good or bad (Horgan & Timmons, 2008).

Still, I find a basic understanding of phenomenological ideas critical to put Coeckelbergh's (2012) social-relational argument into context. A brief review of phenomenological and post-phenomenological efforts is therefore helpful to recognise the importance phenomenological insights have for relational accounts. Furthermore, many of the phenomenological concepts offer a direct point of comparison to the modern assumptions of property-based accounts previously discussed.

The two main contrasting aspects that I aim to point out here are how a phenomenological approach conceptualises selfhood and subject-object relations and, secondly, how a phenomenological view conceptualises technology in relation to us.

Phenomenological Insights

Phenomenology, as its own distinct disciplinary field, has been largely attributed to the work of Edmund Husserl (1913) and is most generally defined as a 'study of phenomena' (Smith et al., 2018). More specifically, it is a study of experiences, consciousness and appearances of things in our lived experience (Smith et al., 2018).

Classical phenomenology has specifically positioned itself as an alternative to the empirical sciences: The phenomenological method seeks to overcome a worldview of reality depicted by the modern sciences and proceeds by describing reality rather than analysing it (Merleau-Ponty & Smith, 1962). Where scientific inquiries aim to make factual claims about the objective nature of things, phenomenological inquiry attempts to uncover the structures of meaning in our lived experiences. Rather than studying what things are in themselves, phenomenology studies the ways in which things come to light, the ways in which a thing shows itself or how things are experienced by us. At the centre of the phenomenological inquiry stands the individual's subjective experience of a phenomenon, or broadly said, the relation between humans and their world.

Phenomenology, overall, can hardly be confined to a single philosophy, which makes an exact characterisation difficult: There are various phenomenological traditions, each

differentiating in emphasis, methodologies and approach. Among them, two of the most significant distinctions are Husserl's (1913) 'transcendental phenomenology' and Heidegger's (1996) 'hermeneutic phenomenology'.

The phenomenological framework I will primarily refer to, and one that is most frequently applied towards our relation to technology, is Heidegger's (1996) hermeneutic understanding. In *Being and Time*, Heidegger (1996) famously analysed our relationship and experience of tools and artefacts with the example of a carpenter using a hammer. His analysis allows us to draw an analogy to how we oftentimes experience technology on a daily basis.

When we encounter technological tools, such as a carpenter picking up a hammer, the tool is put to use without much contemplation. The hammer is almost immediately and readily employed towards hammering and becomes what Heidegger calls 'ready-to-hand' (*zuhanden*; Heidegger, 1996). What is interesting about this example is that it allows us to realise a different view on subject-object relations: By using a tool in a familiar and engaged way, there is no experience of a separation between the alien object and the carpenter's subjectivity – in fact, there is no experience of subjects and objects at all. Instead, 'there is only the experience of the ongoing task (e.g., hammering)' (Wheeler, 2011).

In this way, the self is understood as an experiential, fully embodied subject that is intricately engaged with the world moment-to-moment – contrasting the modernist notion of an objective, observing self. This phenomenological view of self is 'non-dualistic' (Ess, 2015) and establishes a view of 'unitary' or 'integral selves' (Natanson, 1970).

In contrast to the Cartesian separation between mind and body, the phenomenological account roots the experience of self and objects in the individual body itself. In Maurice Natanson's (1970) account, this is referred to as *corporeality*: 'I am neither "in my body" nor "attached to" it; it does not belong to me or go along with me. I am my body (...) I am corporeality' (Natanson, 1970, p. 11). Similar rejections of the Cartesian dualism have been famously articulated in the work of Merleau-Ponty (Merleau-Ponty & Smith, 1962).

The significance of a non-dualistic view of self is elaborated on in an analysis by Charles Ess (2015): By viewing self in an embodied non-dualistic way, phenomenology drastically refutes the modernist idea of subjective vs. objective knowledge and instead gives epistemological legitimacy to subjective first-person experience (Ess, 2015, p. 92). The split between appearances and truths and the resulting epistemological uncertainty that is prevalent in property-based accounts does not apply to a phenomenological or relational understanding.

Within the philosophy of technology, the non-dualistic, phenomenological understanding of selfhood is found in numerous accounts (including Coeckelbergh's social-relational theory). A prominent example is the figure of the "cyborg" found in Donna Haraway's *Cyborg Manifesto* (Haraway, 1991), as well as Andy Clark's *Computational Theory of the Mind* (Clark, 2014).

Haraway's (1991) cyborg describes a hybrid of machine and organism that we can find embodied all throughout the modern world today. Modern medicine is one such example: Here, we already find the human organism modified through intelligent technologies such as microchips or prostheses. The figure of the cyborg human represents a breakdown of the strong nature-machine dichotomy and is applied by Haraway to 'suggest a way out of the maze of dualisms in which we have explained our bodies and our tools to ourselves' (Haraway, 1991, pp. 149-181).

Andy Clark's account articulates a similar understanding: By experiencing things outside of our bodily boundaries – from tools to books to websites – as an integral part of cognition itself, we are already 'natural born cyborgs' (Clark, 2014).

Understanding the phenomenological view of self is important as it plays a prominent role in the social-relational account: Coeckelbergh (2012) encourages us to apply a "Heideggerian" hermeneutic perspective to attend to the ways in which entities are experienced and encountered (Coeckelbergh, 2012, p. 25). The emphasis hereby lays on the encounter as embodied and social beings. There is no possibility for an objectivist, detached view of technologies and their moral status, as we are always already involved and engaged within the world. This further implies that there can be no boundaries drawn between ourselves, the environment or the entities in it (Coeckelbergh, 2012, p. 26).

With the help of these phenomenological insights, we can contextualise the dynamics of human technologies relations today: Not only do modern artificial tools extend our selfhood in the way a hammer does, but they also modify, influence and transform us in turn. In Marshall McLuhan's *Medium theory*, he describes that in our use of technology, a person is 'perpetually modified by it and in turn finds ever new ways of modifying his technology' (McLuhan, 1994, p.46).

The phenomenological understanding of technology offers something that a modernist or property-based view of entities couldn't: It allows for an accurate depiction of some of the modern technological phenomena presently occurring. Rather than upholding a strict boundary between technological entities and human observers, it makes it possible to acknowledge the many ways in which technologies are already intrinsically interwoven with

our own selfhood – whether that is in the case of medical devices, avatars or in personalised algorithms.

More recently, the so-called *post-phenomenological* approach (Verbeek, 2005) has emerged from traditional phenomenological perspectives. It is applied by scholars like Verbeek (2005) as a variant of phenomenology without transcendental pretensions. In addition to overcoming the subject-object split, post-phenomenology reconceptualises the relations we have to technology as *mediation*.

One of the most famous post-phenomenological analyses was done by Don Ihde (1990), who emphasises the mediated character of any subject-object relation: As technologies mediate our perception and relation to the world, there is no direct relation between subject and object, between human and world or between human and technology. Instead, there is only a mediated one. This understanding offers an interesting point of comparison to Coeckelbergh's (2012) account, which will be revisited later.

From Ihde's (1990) and Verbeek's (2005) post-phenomenological perspective, the moral status of technologies is ascribed from a place of 'human-technology associations'. Both entities shape and influence each other in a "worldly" context and are intrinsically intertwined – thereby making it counterintuitive to question their moral status as separated, individual subjects (Feenberg, 2010).

The post-phenomenological approach appears to be even more "radically relational" in the sense that it understands subject and objects as constituted *by each other* rather than understanding relations in terms of intentionality (e.g., being conscious of something, subjectivity directed towards an object; Drummond, 2015). In contrast to phenomenology, in which experiences are understood in the context of some lifeworld, post-phenomenology rejects the assumption that there is some kind of pre-given world of subjects and objects. Instead, it identifies the mediation of technology as its source. Technological entities would not just be passive objects that we relate to but actively shape the ways in which the relation is experienced (Verbeek, 2005, p. 16).

With the phenomenological and post-phenomenological concepts in mind, I will now discuss the social-relational framework provided by Coeckelbergh (2012). As it will turn out, many elements in Coeckelbergh's (2012) account closely resemble the phenomenological understandings of technology. However, it remains questionable whether Coeckelbergh's account truly overcomes the subject/object split in the same way that a phenomenological view does.

Coeckelberghs Social-Relational Account

Coeckelbergh has expanded on his relational view of morality in a number of papers, including various proposals from a ‘relational turn’ (Coeckelbergh, 2010, 2014, 2020) to a ‘linguistic turn’ (Coeckelbergh, 2011). The most comprehensive account of Coeckelbergh's relational arguments, however, was developed in his book *Growing Moral Relations* (2012) and will serve as my primary point of reference.

At the centre of the relational turn is the idea that moral status cannot be determined through metaphysical properties but instead should be based on the way an entity is experienced by us: ‘The main idea is that moral status is not a matter of the properties of the entity (relational or not) but has to do with the way the entity appears to us’ (Coeckelbergh, 2012, p. 214). In this way, Coeckelbergh rejects the possibility of an objectively true and experience-independent answer to the question of moral status.

In the social-relational account, there are no intrinsic and measurable properties that constitute a fact of the matter but only *extrinsic* ones (Coeckelbergh, 2012, p. 195). The terminology of extrinsic versus intrinsic properties that Coeckelbergh applies, however, is one that I prefer to avoid as it can lead to highly problematic interpretations. It commonly induces a misunderstanding of relational accounts as simply shifting from intrinsic to extrinsic properties, thereby still being property-based. This point will be more extensively discussed in a later section.

Similar to Ingold's (1996) and Latour's (2005) accounts and in line with phenomenological traditions, Coeckelbergh (2012) argues that relations are prior to moral status ascriptions: ‘The entity already appears in a particular way, is already interpreted (receptive aspect) and constructed (active aspect) before we ascribe moral and ontological status’ (Coeckelbergh, 2012, p. 88). Although not directly referenced in Coeckelbergh's texts, the idea of relations being prior to moral status is remarkably similar to Levinas' (1979) famous claim that ‘ethics precede ontology’.

To describe the variety of ways an entity can appear in different circumstances, Coeckelbergh uses the post-phenomenological idiom of *multistability* introduced by Don Ihde (1990). Moral status is not fixed for a class of entities but can be dual or multistable depending on *appearance in context*. In one relational context, a robot is simply an efficient tool; in another, the same robot might appear as a communication partner and friend (Coeckelbergh, 2012, p. 48).

The idea of a context-dependent view of entities is something we can find in Floridi's (1999) and Paul Verbeek's (2005) accounts as well: A person can be perceived as a unique individual named Mary, as a human being, as a mother, as a physical body and so forth – depending on the relational context or in Floridi's case, depending on the 'level of abstraction' (LoA) (Floridi, 2013, p. 104).

Another example of the multistability of entities is elaborated on in Verbeek's (2005) example of ultrasound technology. The ultrasound sonogram allows for different ontological statuses of the fetus: For one; it makes it possible to view the fetus as a separate living being and as an individual person rather than only perceiving it in unity with its mother. At the same time, it facilitates a view of the fetus as a patient that can be screened for any medical abnormalities. The technology of the ultrasound, in this example, enables a multitude of possible relations to the fetus (e.g., as a medical subject, an individual person, or a child; Verbeek, 2005, p. 25).

While similar views have been articulated in several accounts, Coeckelbergh (2012) actively seeks to differentiate himself from other relational theories. In Coeckelbergh's account, several relational accounts are reviewed but ultimately rejected for not being 'truly relational' (Coeckelbergh, 2012, p. 195). Examples include Floridi's information ontology (1999), communitarian accounts (Taylor, 1989), ecologist theories and Latour's (1993) actor-network theory.

Communitarian views, such as those found in Aristotle's *Politics* (1984) or Taylor's (1989) account, describe the human as a member of a larger collective body. They allow for a holistic view of relations between members and their community, thereby moving beyond an individual-society dualism.

Still, Coeckelbergh criticises them for being property-based and anthropocentric, as they only apply to relations between human beings. He further argues that they pose a risk of becoming collectivist in practice: 'It (communitarianism) tends to see members of the community as mere means to the communal end' (Coeckelbergh, 2012, p. 37).

Ecologist theories are criticised by Coeckelbergh in the same aspect: 'It (ecology as science) prioritises the whole over parts, natural ecology is collectivist' (Coeckelbergh, 2012, p. 48).

Surprisingly, even Latour's (1993) network ontology and the idea of a 'collective of hybrids' is criticised by Coeckelbergh (2012) through a similar argument and is found to place too much emphasis on the collective: 'since Latour focuses on the collective, it remains unclear how relational his account is' (Coeckelbergh, 2012, p. 62).

That is despite the clearly relational elements in Latour's (1993) account, which view moral status within a network of humans and non-humans (natural entities as well as objects) and reject a separation of the natural and social. What Coeckelbergh finds to be the problem is that this account analyses nature-culture from an external, somewhat "bird-eye" point of view. By doing so, he argues that Latour (1993) takes on a scientific approach rather than a philosophical one and lacks a relational view that is *engaged* (Coeckelbergh, 2012, p. 62).

What exactly Coeckelbergh (2012) means with a more 'engaged' philosophy and a 'truly relational' view finally becomes apparent in his discussion of Ingold's (1996) anthropology. Similarly to Latour (1993), Ingold (1996) aims to go beyond the nature/social dichotomy but does so with a more phenomenological focus (Coeckelbergh, 2012, p. 63). Ingold's (1996) understanding of humans as 'involved beings' and his characterisation of 'knowledge as a skill' developed through engagement with the world seems to be the aspect that Coeckelbergh (2012) is looking for in a truly relational account.

How an entity appears to us depends on the different ways of engagement with the world, as well as on pre-existing knowledge and practices. Coeckelbergh describes that 'both non-scientific and scientific knowledge and practices presuppose lived experience and relations with other entities' (Coeckelbergh, 2012, p. 68).

Coeckelberghs' Conditions of Possibility

An essential element in Coeckelbergh's account, which differentiates his arguments from those of post-phenomenological authors, is his transcendental approach to moral status. Rather than "only" constructing a relational view of the world, Coeckelbergh seeks to ask a transcendental question about the conditions of possibility for moral status ascription (Coeckelbergh, 2012, p. 5).

Conditions of possibility are what make a particular value or status ascription possible in the first place. Coeckelbergh defines and discusses such across six dimensions: 'linguistic, social-cultural, natural-bodily, material-technological, religious-spiritual and historical-spatial' (Coeckelbergh, 2012, p.5).

These conditions of possibility shape and limit how an entity appears to us in the first place: Our perception of an entity depends both on the historical context we are immersed in, our social and cultural background, the language we use to refer to the entity and so forth. For example, the very practice of factory farming and commoditisation of animals is a condition of possibility that allows us to perceive and treat animals as raw materials (Coeckelbergh, 2012, p. 61).

However, Coeckelbergh does not identify these dimensions as *causes* of moral status ascription and rejects any deterministic accusations. The social-cultural dimension does not cause an entity to appear a certain way – it is merely *made possible* by it: ‘Moral status is made possible by a culture and sociality that are not represented or constructed but that are given from the start’ (Coeckelbergh, 2012, p. 8). They exist as a condition prior to moral status ascription and form a shared background that pre-structures our experience. They both enable and limit the ways an entity can appear to us.

The idea of a “shared background” is similarly found in various concepts, including Heidegger’s (1996) ‘being-in-the-world’, Habermas’ (1984) concept of ‘lifeworlds’, or Gier’s (1980) ‘inherited background’.

While Coeckelbergh (2012) briefly mentions these terms, he predominantly refers to Wittgenstein’s ‘forms of life’ (*Lebensform*; Wittgenstein, 1953, p. 238). There are various known and competing interpretations of the so-called forms of life, yet the relevant one for Coeckelbergh is the social-cultural interpretation offered by Peter Winch (1958). From the social-cultural perspective, forms of life are understood as inherently tied to society: ‘Principles, precepts, definitions, formulae – all derive their sense from the context of human social activity in which they are applied’ (Winch, 1958, p. 57). In line with this interpretation, Coeckelbergh understands society and social practices as the ontological source for background familiarity and forms of life (Coeckelbergh, 2012, p. 107).

This understanding appears intuitive: By growing up in a particular culture or society, we learn the status of entities in different contexts and, over time, build a so-called ‘know-how’ and ‘skill’ of moral status ascription (Coeckelbergh, 2012, p. 107). In our daily encounters with entities, we do not actively question the moral status of every entity but rather have a ‘readiness of dealing with them’ (Dreyfus, 1990, p. 149).

The emphasis Coeckelbergh (2012) places on a social interpretation of moral status also explains why he does not only speak of a relational view but suggests a *social-relational* one. Where a property-based account thinks of moral status as a measurable property or feature, Coeckelbergh (2012, p. 107) advocates for a view of moral status as inherently social:

Ascribing status to entities is ‘social’ in the sense that entities (human or non-human) can only appear to me as particular entities based on a background familiarity that is inherently social. In other words, it is because I am already socialised in a world – a social world – that I can talk about the moral and ontological status of entities.

If we were now to ask if intelligent technologies are moral patients in Coeckelbergh's account, there would be no definite answer. Moral status would not be fixed and determined but instead an ongoing activity resulting from our relationship and engagement with entities in the context of a specific form of life (Coeckelbergh, 2012, p. 93).

In the social-relational account, there is no concrete answer to which entities are morally considerable and which ones' moral patiency is ascribed to. Coeckelbergh (2012) himself points out that 'we still do not know what the moral status of particular animals, robots and other entities is' (Coeckelbergh, 2012, p. 195).

Consequently, we can imagine that a robot can appear as an object for one person and be given no moral status but can also appear as a friend to another person – and then have moral status? Coeckelbergh's (2012) account remains somewhat unclear about which appearances actually lead to an ascription of moral status. An underlying notion that persists, however – and one that I argue to be highly problematic – is that the appearance of something as a lifeless object or tool equates to a lack of moral status, whereas the appearance of a human other or friend indicates possession of moral status (Coeckelbergh, 2012, p. 44).

I will challenge this notion and other aspects of Coeckelbergh's (2012) account in the next section and instead argue that our engagement and perception of something as an object is not a reasonable ground to dismiss its moral status.

The Social-Relational Account Discussed

There are a couple of noteworthy advantages that Coeckelbergh's (2012) social-relational account has to a property-based strategy. For one, it offers a better, non-dualistic conceptualisation of technology that closely resembles real-world dynamics. Technology is no longer located in a separate sphere but is understood as fundamentally intertwined with the natural and social. Coeckelbergh emphasises this aspect during multiple points in his analysis: 'Technology cannot be sharply distinguished from society or from "nature"; it is embedded in the relational, social world which consists of both humans and non-humans' (Coeckelbergh, 2012, p. 153). In a relational view, the problem of defining entities within the boundaries of their properties no longer persists. Instead, the social-relation view allows us to conceptualise technological entities in a way that accounts for many of the current technological phenomena we experience (e.g., hybrid entities, a blurring of technical/social/natural worlds, etc.)

While property-based accounts are detached from the reality of moral status ascription, Coeckelbergh's (2012) social-relational argument achieves the exact opposite. By

referring to concepts such as ‘inherited backgrounds’ (Gier, 1980) and ‘forms of life’ (Wittgenstein, 1953), Coeckelbergh is able to describe moral status ascription as close to our lived reality as it gets.

Considering various social-cultural and historical conditions explains the repeated failure to include entities in the moral circle despite them having all the relevant properties. From a social-relational perspective, it becomes perfectly obvious why entities like animals often have conflicting moral statuses and can appear both as lifeless products and as conscious companions: In our forms of life, we have learned that eating animals is socially accepted and, therefore do not perceive the act as immoral or perceive a piece of meat as a suffering moral patient. Our ‘inherited backgrounds’ structure and enable how entities appear to us and equip us with competencies in dealing with them (Coeckelbergh, 2012, p. 185).

In the example of robots, this means that linguistic dimensions of how we talk about robots (e.g., machines vs. companions) or social dimensions of how we live together all ultimately shape how a robot appears to us. It explains why robots can be anthropomorphised in some situations and perceived as a dystopian threat in others.

The social-relational view succeeds in that it can account for the often conflicting perceptions of the same entity in a way that property-based accounts cannot. Furthermore, it offers an explanation for the historical developments of moral status in which humans could appear as ‘qua man’ and ‘qua slave’ (Aristotle, 1984) despite having the same properties.

By shifting the focus from properties to relations, Coeckelbergh’s account allows moral status to quite literally “grow” over time. (Tollon et.al., 2021).

But just because the social-relational account is closer to our lived reality of moral status ascription and offers a more dynamic, historicised account of moral status, it is not necessarily the more suitable framework. In fact, the extensive explanations about the conditions of possibility for moral status ascriptions, which take up a majority of Coeckelbergh’s (2012) book, have led some scholars to criticise his account as being merely ‘descriptivist’ (Nyholm, 2023).

Coeckelbergh’s (2012) account succeeds in giving us insightful answers to questions like “Why do some entities appear to have a certain moral status?” or “What are the conditions for moral status ascriptions?”. However, it does not necessarily give an answer to what the moral status of technological entities actually is.

Similarly, the scholars Tannenbaum and Jaworska point out that relational arguments like Coeckelbergh’s (2012) ‘do not truly offer an account of moral status, but only of particular agents reasons vis- à vis the individual at issue’ (Tannenbaum & Jaworska, 2018, p.

9). It merely allows us to account for different perceptions of moral status in different situations, hereby portraying the concept of moral status as something variable and unstable.

While conceptualising moral status as “multistable” helps explain our relations to technology at a descriptive level, it is simultaneously a point of weakness and a source of frequent criticism. At this point, I will address the two most frequently voiced points of criticism against Coeckelbergh’s (2012) relational account before adding my own concerns. These comprise claims that:

1. The social-relational account is indirectly property-based (Sætra, 2021)
2. The social-relational account poses problems of extreme relativism

The first point has been most extensively articulated in a critique by Henrik Sætra (2021). It expresses the concern that a relational account only shifts from intrinsic to extrinsic properties – yet moral status ascription takes place through properties nonetheless. Despite Coeckelbergh’s claim of being truly relational, it appears that his theory cannot avoid a reliance on properties and is just “more of the same”. This impression seems especially justified when one comes across paragraphs such as the following:

The alternative approach I propose attempts to avoid the scepticism by replacing the requirement that we have certain knowledge about actual ontological features of the entity with the requirement that we experience the features of the entity as they appear to us in the context of the concrete human-robot relation (Coeckelbergh, 2010, p. 213)

Here, Coeckelbergh himself speaks of replacing one requirement with the other – namely, intrinsic properties that constitute a factual, objective reality with extrinsic properties as we perceive them. Properties and features still matter, only now the problem of epistemic uncertainty is eliminated: The properties no longer reflect a true reality but are given a different status, namely as ‘apparent features, features-as-experienced-by-us’ (Coeckelbergh, 2010, p. 213). From this, it seems that the social-relational approach is not as radically different as it claims to be and is relying on properties nonetheless.

However, the difference is that in the case of a social-relational account, the correct criteria for moral status are not predetermined – if it is even determined at all. Unlike the property-based approach, Coeckelbergh (2010) does not identify *which* specific property is relevant for moral status. At no point does he make the argument that an entity matters morally because it appears conscious or intelligent, as it is the case in typical property-based account. What exactly it is that matters in our relation or perception of an entity remains rather vague and unspecified in Coeckelbergh’s case.

One might say that “having an interpersonal relation to entity X” or “feeling empathy” about an entity is just another kind of property or condition for moral status ascription. In this case, the status ascription appears to be even more susceptible to anthropocentrism than a property-based account. Even though Coeckelbergh claims that his argument is neither rooted in the subject nor the object, it does seem like, ultimately, only the human’s perception of an entity matters. Sætra (2021) similarly describes this and claims that the relational turn is not a move away from anthropocentrism but instead a new-found *neo-anthropocentrism*. In Coeckelbergh’s argument, moral status is deeply dependent on what or who humans perceive to have value: ‘It is human perceptions and experience of the other that is used as the basis for determining value’ (Sætra, 2021, p. 7). The approach remains heavily one-sided, namely on the side of the humans, as only what we see and perceive to matter has relevance (Sætra, 2021).

While I similarly find the relational approach too one-sided, it is important to note that moral status ascriptions are not meant to be understood as solely about human perception. Commonly, scholars misinterpret Coeckelbergh’s argument as something along the lines of “If something *appears* to have x property, then it has moral status” or, as Müllers (2021) phrased it in his discussion of Coeckelbergh’s account, ‘If we care about something it has moral status’.

However, an interpretation of Coeckelbergh’s account in which moral status ascriptions are solely a matter of perception would do it injustice. In his concluding remarks, Coeckelbergh explicitly defends suppositions of this kind and claims that his work goes further than post-Kantian positions and shows that moral status ascription is ‘not like the perception of a property and not like perception at all’ (Coeckelbergh, 2012, p. 207). Instead, he claims that moral status ascription, in his understanding, is a matter of active *doing* as opposed to *seeing*: ‘Moral status ascription depends on how the subject engages with the world’ (Coeckelbergh, 2012, p. 108).

Yet, even if this is taken into account, it is hard to imagine how the similarities to the anthropocentrism in property-based accounts can be overcome. Even if relations are understood as “active” and “engaged” doing, how we relate to an entity still ultimately depends on its properties. Those entities with more human-like properties (such as humanoid robots or chatbots) will more likely induce an engagement that is similar to our relations with human others. While properties and characteristics may not be solely responsible for how we relate to an entity (as this is still highly context-dependent and influenced by various social and cultural factors), properties continue to play a significant role.

As a result, the relational approach generally ends up providing moral status to the exact same entities as before and will likely resemble the outcome of a property-based account: Animals and other humans have a higher chance of moral status ascription, whereas technological entities would only receive moral status in isolated instances. Overall, the relational account appears to actually be ‘a camouflaged variant of the property-based approach’ (Sætra, 2021, p. 7).

Characterising the social-relational account to be entirely the same as a property-based one might be a bit too radical at this point. After all, Coeckelbergh’s (2012) account is already different in that it establishes transcendental conditions for moral status ascriptions and analyses the self and other entities in a phenomenological way. What I find to be a more justified criticism of the account is that moral status ascription ultimately remains relativistic and human-centred.

The Dangers of Relativism

The most frequently voiced criticism against Coeckelbergh and other relational accounts is that they impose problems of relativism. Tollon and Naidoo (2021), for example, state that the social-relational argument has a ‘problem of moral relativism’ (Tollon & Naidoo, 2021), and others even express a fear of ‘extreme meta-ethical relativism’ (Moser & Carson, 2001). The term relativism encompasses many different ideas and variations and remains widely controversial among philosophers.^{ix}

Claiming that an argument faces a problem of relativism is entirely different from claiming that it has a problem of *moral relativism*. The idea behind relativism overall is not that every feature of an entity is ultimately unknown and merely up to decision (e.g., we don’t know what colour a robot is (green, blue, grey) because it is relative to our perception). It does not call for us to reject the possibility of scientific facts as a foundation of knowledge.

Relativism more so embodies efforts within the so-called Counter Enlightenment movement which question how human thought and conceptions are shaped in the first place (Swoyer, 2010). The point various relativistic thinkers generally tried to express is that there is no single account of truth and reality, and in ‘only a perspective seeing, only a perspective knowing’ (Nietzsche, 1966).

Moral relativism, on the other hand, is concerned with the relativity of morality rather than truth and reality. Metaethical moral relativism, specifically, can be defined as the idea that there is no universally valid or correct moral judgment. It holds that the truth or falsity of

moral judgements is not absolute but always relative to the convictions of a group of persons (Gowans, 2021).

Coeckelbergh's (2012) relational account can be said to be relativist in the sense that there is no objective, universal truth about the moral status of an entity: In one context, a technological entity might be a moral patient, and in another, it has no moral status at all.

As a result, the relational approach is frequently accused of facilitating an ethics of 'anything goes' (Müller, 2021). Scholars like Müller (2021) have claimed that Coeckelbergh's (2012) argument allows us to take moral status to absurd levels, in which even a pencil can have moral status simply by means of a human caring and feeling attached to it (Müller, 2021, p. 581). And caring obviously seems to be an extremely subjective standard for moral status ascription and hardly constitutes valid grounds for justification. As Gunkel has pointed out (2012) with reference to Benford and Malartre (2007, p. 163):

We don't care about a million people dying in China of an earthquake, ultimately, in an emotional way. We try to, but we can't really because we don't share the same physical space. It might be much more important for us if our dog is sick.

What entities we care about might be one of the most subjective standards of all. Similarly, Mosakas (2020) criticised the social-relational approach as being too 'fundamentally concerned with our feelings and attitudes towards different entities' (Mosakas, 2020, p. 431).

The reliance on human perception in the social-relational account yields deeply problematic consequences due to our own human nature. It is well known that the human mind falls victim to a number of psychological fallacies and is inherently irrational (Kahneman et al., 1982). There are a number of situations in which even our fellow human others are objectified and thereby potentially become excluded from moral status – especially when the encounter with another human subject is not physical but instead takes place in virtual spaces. When the human other is encountered only as an avatar in a video game or social-media username, studies reveal increased levels of dehumanisation, decreased empathy and prosocial behaviour and a generally lower perception of moral significance (Anderson et al., 2010; Bandura, 1999; Greitemeyer & Mügge, 2014). In some situations, we might find a much higher moral status being attributed to non-human entities than other human agents. This is especially solidified by studies showing that humans have a tendency to anthropomorphise and empathise with certain technological objects, such as robots (Złotowski et al., 2015). In some extreme cases, they may even become romantically attached and form emotional connections to inanimate objects (Garber, 2013).

What conclusion does an account like Coeckelberghs lead us to? Should we just accept that the moral status of technological entities is multistable –sometimes attributed, sometimes not – depending on its relational context? And if so, is the morality of our actions nothing more than a matter of personal perception and judgment?

Again, the word “perception” must be applied with caution. Claims that moral status in Coeckelbergh’s (2012) account is solely a matter of perception or as ‘humans caring about things’ (Müller, 2021, p. 581) would misconstrue the social-relational argument.

Although Coeckelbergh (2012) himself speaks of ‘appearances’ and ‘perceptions’, he also warns that this kind of wording poses the risk of a false understanding in which moral status ascriptions are made from an outside view and by a disengaged observer. But as subjects that are “already-in-the-world”, moral status ascriptions are always active: ‘What makes moral status ascription possible is not a matter of seeing, but doing’ (Coeckelbergh, 2012, p. 207).

Yet even if we replace “seeing” with “doing”, and even if moral status is not a result of appearance but of engagement – the problem of relativism still persists.

Of course, ‘anything goes’ cannot be taken in a literal sense, as there are still limits and boundaries to how we engage with entities. As Coeckelbergh states in his defence: ‘Not all meanings and perceptions are possible (you cannot make a cow from the vase)’ (Coeckelbergh, 2012, p. 81).

Further, it can be argued that our engagement with an entity is constrained by the entity’s properties. We might be able to talk to a chair, but we cannot communicate with it, as communication requires an interactive exchange from both sides.

But while our engagement is not completely limitless and is influenced by cultural and historical factors, societal norms and so forth, it is still highly subjective to the individual. It can take extremes in which the same class of entities is engaged with as a product and means to an end in one situation and as a best friend in another.

While I share the worries scholars have about the extreme variability of moral status in the social-relational account, it would be false to characterise Coeckelbergh's relativism as a *moral relativism*. The relativism of the social-relational argument consists of an opposition against objective knowledge and empirical truths. It is meant to challenge our beliefs about the existence of certain properties (e.g., consciousness, sentience) by emphasising the embodied aspect of our human experience.

What the social-relational account does *not* do, is claiming that the morality of our actions is relative and that there is no “right” or “wrong”. While Coeckelbergh's account

potentially poses a risk of moral relativism, it never actually makes a case for such. Quite on the contrary, Coeckelbergh (2012) warns about prospects of moral relativism and advocates for a so-called ‘ethics of skill’, which I will briefly review in the following.

In Defense of Moral Relativism Accusations: Proposals of an Ethics of Skill

In defence against accusations of relativism, Coeckelbergh put forward the idea of a so-called *ethics of skill*. Rather than claiming that no right or wrong moral judgements are possible, Coeckelbergh argues that there are some ways of relating to an entity that are ‘better than others’ (Coeckelbergh, 2012, p. 155).

However, finding out what a “good” way of engaging with an entity is is not an easy process. According to Coeckelbergh, the answer to what kind of moral relations we *should* have to entities cannot be decided a priori or from a ‘point of nowhere’ (Coeckelbergh, 2012, p. 155). Instead, it needs to be actively discovered through an experimental approach and an ethics of skill.

Inspired by proposals of *experimental virtue ethics* in various environmental accounts (e.g., Cafaro, 2004; Hursthouse, 2007), Coeckelbergh argues that we cannot know virtues a priori, but have to form them in practice and based on experience: ‘True moral knowledge is in the skill and activity in which good is experienced and done’ (Coeckelbergh, 2012, p. 163). An example Coeckelbergh (2012, p. 163) uses to demonstrate this is the moral status of animals:

If we have never slaughtered an animal, we do not really know what we are doing when we support animal factories by eating the meat they produce or when we write about killing animals. And, if we want to know the good life, we have to experiment, try out different possibilities.

Apparently, the experience of the animal being slaughtered would likely cause us to be miserable: ‘We would feel that it was not right’ (Coeckelbergh, 2012, p. 28).

The idea is that actively engaging and experiencing entities around us will allow us to gain moral experience about what is good and bad. From this, Coeckelbergh (2012, p.35) concludes that the social-relational argument is not relativist because it is always constraint and shaped by our practical moral experience:

Even in the absence of absolute moral standards, moral experience teaches us that some ways are not good for people and their societies. (...) In the case of animals, this could mean pointing out that we learned to respect (at least some) animals more than we did in the past and that this now regulates our behaviour towards animals.

It is important to keep in mind that our behaviour and ways of doing things are deeply embedded in our *form of life* and *habitus*. How we behave towards entities is generally not an individual act of will but is more so an adaptation to societal practices around us.

Over time, certain ways of relating to an entity are increasingly reinforced and ingrained in our habits and “know-how”. Having such know-how is practical in that we do not need to debate every single course of action but instinctively know how to react. In our daily lives, we do not need to question whether setting a car on fire is morally wrong – we already intuitively know so. However, this also means that the likelihood of a sudden change in our behaviour and perception of the moral status in technological entities is low.

Coeckelbergh acknowledges this difficulty as well: ‘Developing a new habitus is difficult. (...) changing views about moral status requires changing society, changing cultures, changing forms of life’ (Coeckelbergh, 2012, p. 124). Adopting a different view on the moral status of technologies and a different way of doing, cannot not be done as an individual act of will alone and certainly not in a short time frame. Instead, it can only grow within a culture over time (hence the title *Growing moral relations* of Coeckelbergh’s book).

While Coeckelbergh’s (2012) proposal of an experimental virtue ethics might be successful in the case of animals or nature, I find it to be highly problematic in the case of intelligent technologies. Overall, it makes for a rather weak defence against relativism and faces at least three problems that I will address along the following:

1. An experimental approach is only possible on rare occasions and only towards a small scope of entities
2. In the cases that it is possible, it is unlikely to affect moral status ascriptions
3. It continues to neglect the moral status of intelligent technologies

Implications of an Ethics of Skill for the Moral Status of Technology

The most obvious opposition that someone might voice in response to Coeckelbergh's proposal of an ethics of skill is that it is far from being practically applicable. It is unlikely that the average individual, let alone humanity overall, is ever inspired to visit a slaughterhouse and thereby achieve a different perception of the moral status of animals.

Our perception of what is morally right or wrong largely depends on the societal and cultural context around us – hence why we don’t naturally find the purchase of meat to be morally problematic. Similarly, we wouldn’t typically find the act of buying a t-shirt from a fast fashion retailer to be immoral, despite there being good moral reasons against it. The likelihood of someone visiting a clothing manufacturer to realise the morally questionable

implications of their consumerism would be just as low. If experiencing things first-hand is how we discover what is morally “good”, then attaining Coeckelbergh’s idea of ‘true moral knowledge’ seems highly unlikely in practice (Coeckelbergh, 2012, p. 153).

On the other hand, participation in experimental virtue ethics does not necessarily require extreme measures: In Cafaro’s (2004) analysis of Thoreau’s Living ethics, he claims that something as simple as going outside and sitting in the forest can be enough to form a ‘joyful relationship’ with nature and can induce a different moral status perception of natural entities.

Yet even if we assume that people *are* able to engage in experimental virtue ethics – that it is somehow possible to experience animals or products in a different way – the problem of moral relativism does not disappear. There is no reason to be sure that all humans will share the same intuition about “right” and “wrong” actions when confronted with a certain entity. Some might experience empathy and a new appreciation for animals after visiting the slaughterhouse, while others might remain unaffected. And even if a prior perception of moral status is challenged, the impact of the experience and the impression might be short-lived. In most cases, we will fall back to our usual habitus and deeply ingrained moral “know-how”.

Overall, I find that Coeckelbergh overestimates the human ability to determine what is “morally good” a bit too optimistically and underestimates the role of cognitive dispositions and fallacies. What we perceive as morally right or wrong is a highly complex process and is constrained by a number of biases.

In a paper by psychologist Kurt Gray (2012), a variety of empirical studies are introduced which show that we tend to rationalise our own misconducts and behaviours to reduce *cognitive dissonance*. In psychology, cognitive dissonance describes the feeling of discomfort that occurs when a person holds two contradictory beliefs.

In moral situations, cognitive dissonance can cause us to minimise the immorality of our actions. If we knowingly harm an entity, for example, we typically reduce the properties we ascribe to them in order to sustain a positive perception of ourselves as virtuous humans. An example of this phenomenon is found in a study by Loughnan et al. (2010), in which two groups of participants were shown a video of a “suffering” cow. In one group, participants were given a piece of beef jerky to eat prior to the video and in the other, participants ate cashews. The group that ate the beef jerky ascribed a significantly lower capacity to suffer to the cow in a follow-up survey and found it to have less or no sentience at all. On the other

hand, the group which consumed cashews ascribed overall higher moral status and properties like mental states to the cow (Loughnan et al., 2010).

The particular phenomenon taking place here can be described as ‘moral disengagement’ or ‘motivated forgetting’ and is certainly not only limited to cows (Bazerman et al., 2011). It is just one example of many which demonstrate how unreliable and easily manipulated human perception is. Our interpretation and experience of the world around us fall victim to numerous other fallacies and cognitive biases, which have been extensively explored in psychological research (Thomas, 2018). Relying on humans' moment-to-moment reactions to determine moral status based on what “feels right” appears to be a flawed strategy in most cases.

The implications of effects like cognitive dissonance become especially evident in the case of social media or search engine algorithms: Generally, humans don’t want to be confronted with content that challenges their own belief systems. We don’t want to think of ourselves as bad or immoral persons – which is why we often find ourselves stuck in our own personal filter bubbles.

It is also why certain content or information which exposes the immorality of our own actions is likely to have significantly less algorithmic reach (Pariser, 2011). A passionate wine drinker is much more likely to share a video about the supposed health benefits of red wine with his friends than one who condemns the habit. Of course, wine drinking is not the most morally significant example, however we can imagine a similar scenario in regards to topics such as climate change, social inequalities and so forth.

A video about the impact of short-distance flights on climate change, for example, is less likely to be shared and go “viral” because it challenges our own behaviour and morality. On the other hand, violent or sexist content has been shown to grab our attention much more easily and is consequently shared with a larger audience (Shen et al., 2021).

In our daily actions, we shape the morality of the world around us – now with the means of technology in a way that is faster and more far-reaching than ever. The impact of one Google search or social media “like” might just be marginal and not even morally significant at all. But the idea that one individual’s act is too small to make any difference and is therefore not morally wrong – also known as the claim of individual causal inefficacy (ICI; Sinott-Armstrong, 2005) – is a dangerous one that has been challenged by a number of authors (e.g., Hiller, 2011; Vanderheiden & Steve, 2007).

The resistance against acknowledging the morality of these actions usually stems from a resistance against an attribution of moral responsibility to an individual, when really it

is a much larger collective at work. How we should think about human agency and responsibility as a consequence of this, is a whole different question to be asked.

The point, however, is that our interactions with media and algorithms have at least some moral character. And an ethics of skills, as proposed by Coeckelbergh, will hardly allow us to recognise how much these entities and our interactions with them actually matter.

Complicating matters even further is that an experimental approach to ethics is simply not possible for all the different kinds of intelligent technologies. How would someone experiment with their relation to entities like smart doorbells, intelligent hospital beds or a pair of VR glasses? And what about in the case of ambient technologies that are simultaneously at work in our surroundings? Is it even possible to explore different forms of engagement with ambient technologies when we are barely aware of our relation to them in the first place?

The case of robotic entities is one of the only instances where Coeckelbergh's experimental approach is somewhat realisable. Here, it might be possible to imagine an experiment in which we treat a robot as a friend or human companion in one situation and as a servant or instrument in another. This could also be applied to the example of virtual assistants or chatbots, where one might experiment with different communication styles. Yet, the answer to which way of treating the entity is better than the other is still likely to turn out to be subjective to the individual user.

Another underlying weakness of Coeckelbergh's account that becomes evident from this is that he understands moral status in connection to practical behaviour. In certain parts of his analysis, it becomes clear that the question of moral status is actually pursued to discover how we should engage with entities: 'The question of moral status is really about how to shape that relation and its related practice, activities and experiences' (Coeckelbergh 2012, p. 147). This kind of understanding is based on the same widespread assumption that moral status is a matter of how we treat an entity.

If an entity is treated with respect or if we have some duties towards it – then it is assumed to have some moral status. If it is treated as a means to an end, it is found to lack moral status. In this view, an attribution of moral status is incompatible with an understanding of technological entities as instruments that are used to fulfil certain ends.

However, as I suggested earlier, moral status needs to be thought about separately from our behaviour or an entity's properties. Just because an entity is employed as a tool and as an end to our means, it should not automatically disqualify from moral status.

Thinking of moral status connected to the way we “act” or “treat” entities is based on the supposition that morally considering an entity requires a change in our behaviour. It insinuates that if something matters morally and deserves moral consideration, we have certain moral obligations or duties towards it. However, if we take on the moral status definition as provided by Thomas Birch (1993) earlier, this supposition no longer needs to be held. It might never be necessary to “respect” or proactively “consider” certain technological entities. We may have no reason to change how we act towards intelligent devices and can continue to employ them as instruments serving our needs, and yet we can acknowledge their moral status nonetheless.

In conclusion, it turns out that the social-relational argument for moral status continues to be affected by moral relativism, even when Coeckelbergh’s defence of experimental virtue ethics is considered. The kind of relations we have with entities will still be highly subjective and individualistic, just as much as our experiences of different “ways of doing” will be highly subjective to cultural or societal influences.

The moral status of entities like intelligent artefacts is likely to continue to be underestimated, whereas the status of others, like humanoid robots, is likely to become overestimated or even symmetrical to that of humans. If moral status is thought about in the way Coeckelbergh suggests, a majority of technological entities continue to be left out.

Although Coeckelbergh claims that in his understanding, moral status is neither located in the human subject nor the object but in the relation *between* them (Coeckelbergh, 2012, p. 207), it seems that moral status is actually located in the relation the human subject has *to* the object. Rather than being grounded in the interrelation between humans and technologies, moral status becomes purely grounded in the relations humans have to technology, specifically in the way they act and engage towards them. In the following, I propose to adapt the relational view of moral status in a different way to overcome these anthropocentric tendencies.

Expanding the Relational Argument

The problem I find with Coeckelbergh's argument is not that it puts relations at the centre of moral status – this is, after all, a perspective that I am highly in favour of. It is rather the constrained and one-dimensional way in which relations are viewed in Coeckelbergh's account that I find to be troublesome.

The analysis of relations in the social-relational account is constrained in that it only considers relations in a bilateral constellation from one entity to another (e.g., from subject to

object, from human agent to technological patient). Worse yet, it predominantly focuses on the relations *we* have to technology while neglecting the relation technological entities have to us or facilitate for us. It takes, as pointed out by many other scholars, a deeply anthropocentric view of relations in which only the perspective of the human *relata* matters (Lima et. al., 2021; Sætra, 2021; Tollon et. al., 2021)

In many cases, there is a mismatch between the nature of the relationship as experienced by us and the kind of relation insights and research from Science and technology studies (STS) indicate we have (Sismondo, 2010). The significance of our relations to certain entities (e.g., smart watches, chatbots, social media applications, etc.) is chronically underestimated by us, as we are often not aware of its full scope. Relying on the human relation alone, has proven itself to be a flawed method.

In an attempt to “expand” upon Coeckelbergh’s proposal, I suggest that what needs to be considered are not merely the relations we have to an entity but the *interrelations* (or rather *intra-relations*) facilitated throughout the overall network of entities. Our interaction with one technological entity is rarely confined to the respective entity alone. They do not only facilitate a relation to the specific entity itself but simultaneously facilitate interrelations to other technologies in the broader cybersphere, a relation to other human subjects and even a relation to our individual selfhood.

To illustrate this point, I will draw from the frameworks provided by Bruno Latour (2002), specifically his notion of the *fold*, as well as Barad’s (2007) concept of *intra-action*. This will allow for a realisation of a relational view that goes beyond the borders of the individual entities involved and offers a more accurate depiction of relations today.

Technologies as mediators

Bruno Latour's (2002) concept of the fold has already been extensively discussed for the morality of technologies in Verbeek's (2011) account. However, it was here only applied towards questions of moral agency.

Latour’s (2002) terminology of the fold was introduced in his account on technology and its relations to human beings. The term was specifically chosen to express a view of all moral actants being “folded” together in time and space as they interact with each other (Verbeek 2011, p.45). Each technological entity is understood to have a temporal dimension interwoven into it (e.g., an old hammer embodying the antiquity of the material it is made of) and, therefore, inhabits a specific place in time and space. These temporal elements, as well as the involved entities, are said to become folded and intertwined with each other in their

interaction. This concept might appear abstract and hard to follow at first, yet the idea behind it is simple. Latour's intention behind the concept was to express how both entities (technology and humans) shape and constitute *each other* (Latour, 2002).

For instance, in the case of using a hammer, employing the tool does not merely extend our own bodily functions or merely serves as a means to an end: 'It is impossible here to proceed as if the hammer 'fulfilled a function'' (Latour, 2002, p. 250). Instead, Latour describes how the object of the hammer 'provides for my fist a force, a direction and a disposition' (Latour, 2002, p. 249), which the hand would otherwise not have.

In this way, technology is not simply an *intermediary* in a technical action but a *mediator* which shapes both the human user and the reality around us (Latour, 1996). This is expanded on in another example by Latour, this time with the example of speed bumps on roads. They are not simply neutral instruments, replacing the role of a police officer, but play an active role in shaping the morality of our actions (Latour, 2002, p. 250).

The notion of technology as mediators rather than instruments and their role in constituting the human actant is precisely what Latour's idea of the fold boils down to. Similar to the accounts of Etienne Souriau (1943) and Gilbert Simondon (1989), Latour defines morality and technology as ontological categories, which the human being comes out of but does not stand at the origin of (Latour, 2002, p. 256).

If Latours' conceptualisation of technology is applied, simply speaking of humans "relating" to technological entities does not go far enough. The configuration of technological entities and human entities cannot be viewed as that of isolated beings existing independently of each other, and so their morality cannot be understood to exist within themselves either: 'Nothing, not even the human, is for itself or by itself, but always by other things and for other things' (Latour, 2002, p. 256).

Throughout Latour's philosophy, we can find this same rejection of the subject/object split in a plurality of formulations: 'There exists no relation whatsoever between the 'material' and 'the social world' because it is this very division which is a complete artefact.' (Latour, 2005, p.76). Although Coeckelbergh's social-relational account makes an attempt to overcome the split between the relata, it ultimately falls back on the same dualistic paradigm that we find in property-based accounts.

The aspect in Latour's account that I especially seek to highlight is the ability of technologies to shape ourselves and the world and space around us. Whenever we interact with technological entities, we do not simply interact with them alone but are simultaneously engaging in the shaping of our life worlds and surroundings. In the examples of the hammer

or speedbump, this might not be immediately evident to the reader – however, in the case of modern intelligent technologies, it is ever so obvious.

Machine learning algorithms, for example, are known for their ability to gather information from their users and “feed” on such information to adapt their own behaviour (Esposito, 2022). The outcome of the behaviour of intelligent machines and devices can be partly attributed to the predetermined contingency of the machine itself, as well as individual user input and inputs from other users. ^x

The previous interactions and relations other humans had with the machine partly shape the way it now presents itself to us, and vice versa, affecting the encounters others will have with the entity. This is most evidently seen in the case of Filter Bubbles mentioned earlier, in which we do not just encounter the algorithm itself but an accumulation of values from a social collective of other persons (Pariser, 2011). Technology, in this case, *mediates* the perception of the world around us as well.

Even our own self and identity are replicated within the cyber sphere in the form of so-called data shadows generated from our habits and information shared online (Westin 1968). Every interaction with intelligent technology is, thereby, in some way, a confrontation and interaction with our own selfhood. This idea has been similarly put forward by Friedman (2020), who has argued that human agents are simultaneously the moral patients of their own moral actions.

Intra-relational constellations and systems

Another way to describe this unique moral dynamic within technology relations is to speak of *intra-actions* rather than interactions (Barad, 2007). Karen Barad’s account is primarily concerned with providing an ethical framework for questions of moral agency in socio-technical practices, yet it can be relayed to moral patiency as well.

The term ‘intra-action’ itself is not unique to Barad’s framework and has been applied by numerous authors, including Introna (2007) and Suchman (2007). In any case, it expresses a similar recognition of humans and technological entities constituting each other in their entanglement (Barad, 2007).

One might question why it is beneficial to speak of intra-action rather than interactions. Not only is this distinction advantageous, but it is necessary given the nature of our modern technological surroundings. Merely speaking of interaction between technology and humans falsely insinuates a separation between the entities.

The very meaning of the Latin word ‘inter’, which the term stems from, is translated as ‘between’ or ‘in the midst of’. Speaking of interaction locates relations *between* two pre-existing entities rather than as something *within* them (Draude, 2020). For instance, human and computer relations would be reduced to an exchange of orders and information through the means of the computer’s surface – both entities hereby existing independently from each other. Yet, when considering the role of computers in our daily lives - not to mention the omnipresence of algorithms and ambient technologies described earlier – speaking of interaction alone is not sufficient.

Humans and technologies today oftentimes do not just simply exist in relation to each other but in relation to other systems and entities interconnected with them. We can, therefore, more accurately describe them as existing within entire *relational constellations* made up of a multitude of technical and non-technical entities (Coulton & Lindley, 2019).

Examples like the IoT or smart homes demonstrate the network-like character of such constellations in the most obvious way. Every intra-action with one entity in the system is, at the same time, an intra-action with all other integrated devices. The relations within such technology constellations are constantly being redefined and mutually affected by the entities in them. These dynamics are similarly described in Human-Computer Interaction (HCI) literature, in which scholars characterise the role of the human user as an ‘active, meaning-making individual’ (Höök, 2009, p. 11).

When moral status is only thought of as the way in which we relate to an entity, the actual scope of our relations remains overlooked. Coeckelbergh's social-relational approach can only account for relations as the human observer experiences them, yet it remains ignorant of those relations which are beyond our imminent perception.

Some might object that an intra-relational perspective still does not provide a definite answer to what the moral status of technological entities is and lacks practical guidance. However, what an intra-relational perspective does provide, and what makes it advantageous to Coeckelbergh’s relational account, is the possibility to view relations beyond just a human viewpoint. It allows for a view of an active relating between a system of relata, which naturally opens up wider possibilities for moral status attributions. How these possibilities will be realised and what kind of moral significance is ascribed to the entities as a result is a different question. Yet without a recognition of their intra-relations, the moral considerability of these entities might remain entirely neglected.

Conclusion

In this thesis, I have set out to re-frame the question of moral status towards technologies across multiple different aspects. For one, I have suggested shifting the very starting point of moral status endeavours from questions of duties and responsibilities to a questioning of their moral considerability (Goodpaster, 1978). Further, the conceptualisation of the moral subject itself has been challenged to replace the widespread instrumentalist and atomistic conception of technological entities with a view that allows us to account for their mediating and hybrid character.

At the beginning of this thesis, I posed the question of how we can best think about the moral status of intelligent technologies. Specifically, the question that was found to be relevant is not *how much* we should consider these entities but *what* we should consider about them. In Coeckelbergh's account, the answer to that question lies in the relations we have. *What* is being considered is the way we actively relate and engage with certain entities (Coeckelbergh, 2012). This perspective allows for a significant improvement to the properties approach, where the only aspect worth considering is the characteristics an entity has.

Only relying on properties has been shown to be a problematic approach due to various difficulties in *determining*, *detecting* and *defining* the relevant properties (Gunkel, 2012). I have further analysed the underlying metaphysical assumptions that a property-based view holds, which additionally turned out to be problematic in the face of hybrid entities and ever-increasing overlaps between social and technological spheres.

The relational approach appears as an attractive alternative to overcome exactly those problematics and offers a more integrated view of human and technological entities. Yet, Coeckelbergh's social-relational approach did not prove to be as “relational” as it claims to be and is heavily dependent on subjective human perception and practices. Both standard and non-standard approaches ultimately turn out to be incapable of keeping up with rapid rate at which technological entities and systems are currently evolving.

Although I did not set out to fully develop a new strategy of moral status ascription in this thesis, I have pointed to alternative ways in which the social-relational approach can be reconceptualised. With the help of various relational and phenomenological insights (e.g., Latour, 2005; Barad, 2007; Heidegger, 1996), it is possible to view moral status as something that is neither intrinsically found within an entity's properties nor as something we ascribed to it in relation. Instead, moral status in the form of patiency is realised *within* entities in their intra-relational encounters with one another. While we cannot speak of intelligent

technologies as having moral status *in themselves*, I hope to have made it clear that their morality should not be approached as a matter of individual entities in the first place.

The suggestion to think of moral status as a matter of intra-relations amongst entities should, however, not be taken as a claim that renders properties obsolete. The very way in which novel technologies facilitate relations today is thanks to their dynamic and interactive characteristics. When it comes to questions of moral significance, such as whether technological entities should be ascribed rights or not, questions about their sentience or interests might very well be justified. Properties still matter, yet relying on them as a sole basis for moral status attributions is bound to fail.

Even though my inquiry did not aim to give practical guidance for our behaviour towards intelligent technologies, some practicability still arises. Cultivating a different way of thinking about moral status allows for a heightened sensitivity towards entities that would otherwise remain outside our consideration. While a recognition of these entities might not warrant a change in behaviour, it certainly allows for an increased awareness about the moral significance our interactions with them have. The dynamic nature of human/technology relations poses as much of a difficulty for moral status questions as it makes for an opportunity. It is an opportunity to reconsider 'how to live' (De Vries, 1999) in a world of entangled moral subjects. In spite of the human biases currently dominating ethical theory, my hope remains that we can recognise these opportunities to shape our future relations to intelligent technologies in a more meaningful way.

Endnotes

Heidegger's conceptualisation of technology has evolved over various accounts and cannot be characterized as simply instrumentalist. Quite on the contrary, his analysis in *The Question concerning Technology* famously holds that 'the essence of technology is by no means anything technological' (Heidegger, 1993). For a more extensive review of Heidegger's technology understanding see Blitz (2014) and Dreyfus (1997).

ⁱⁱ Other accounts supporting such view can be found in psychology research and related literature. Specifically interesting here are accounts discussing a so-called negativity bias. By now, various studies have confirmed that we have a human tendency to pay much more attention to negative (or 'low-valence') words and impressions than neutral or positive ones (Carretie, 2001).

ⁱⁱⁱ Initially, phrasing moral status questions in terms of moral importance seems to make for an improvement and appears to be more applicable in the case of non-human entities. However, it turns out to not make that much of a difference after all. Regardless of whether we ask 'Does x have moral importance' or 'Does x have moral status' – the same difficulties occur. In each case, the question remains how exactly such term can be defined, determined and ultimately ascribed to an entity. How does one determine whether an entity has moral importance or not? What are the requirements for something to have so-called 'moral character'? Even though Johnsons (2006) account offers a different, and somewhat unique application of the terminologies, it ends up being just more of the same. Moreover, terms like 'moral character' are ambiguous in nature, which makes their use prone to misunderstandings or false interpretations.

^{iv} One example put forward by O'Neil (1997) to demonstrate why intrinsic value and moral status should be understood separately, is the case of languages. Abstract entities like languages make for an interesting case because they can be of both instrumental and intrinsic value, yet in no case would their moral status be even considered. O'Neil (1997) argues that there might be several reasons why we should keep languages alive, such as that they ought to be preserved for their grammatical and lexical richness or special syntax form – all of which can be considered as intrinsic values of a language. Whether this makes them a sufficient candidate for moral status however, is another question.

^v It is worth mentioning however, that this kind of opposition from Coeckelbergh against indirect arguments is primarily found in his 2012 publication *Growing moral relations: Critique of Moral status ascription*. In a later article, he argues in favor of rights ascription to robots and specifically refers to his arguments as being indirect in the title: 'Four arguments for the indirect moral standing of personal social robots' (Coeckelbergh, 2021). Although Coeckelbergh points out that he applies an indirect justification for moral status in this case, he also emphasizes this by only

speaking of ‘indirect moral standing’ rather than moral status. In contrast, his social-relational account claims to offer a justification for the moral status of technological entities.

^{vi} Despite Birchs strong stance against a property-based approach, which he finds to be ‘a function of imperial power mongering’, he ironically is not able to avoid the use of a criterion himself (Birch 1993). Although Birch suggests a form of ‘universal consideration’ that supposedly takes everything in consideration, it becomes clear in the last chapter of his account, that considering ‘everything’ only refers to the ‘whole biosphere’ (Birch, 1993 p.331). Therefore, even Birchs’ notion of universal consideration, draws a line based on properties and relies on a condition, namely that of being members of the biosphere.

^{vii} When the term ‘dualism’ is applied in philosophical literature, scholars typically refer to a ‘two-world metaphysics’ (Søraker, 2007) meaning a distinction of appearances/reality or to the Cartesian dualism between mind and matter (Descartes, 1637). The way in which a ‘dualistic worldview’ is meant to be understood here however, is based on Søraker (2007) application of the term and simply refers to the separation of entities into distinct classes.

^{viii} Although Danahers behaviorism is used as an example for a relational theory of moral status by some authors like Mueller (2021), it does not represent the understanding of a relational account taken on here. Danahers account does not allow for a shift from properties to relations, but only from intrinsic properties to extrinsic ones. In this case the condition for moral status is simply whether an entity is able to exhibit certain behaviors such as speaking or walking. Drawing on my definition of indirect and direct moral status earlier, Danahers (2020) account only allows for an indirect status ascription and is not “truly” relational.

^{ix} For a better overview of the variations and scope of relativism (‘global and local’ ‘strong and weak’ forms of relativism or epistemic relativism) see Swoyer (2010) and Köbel (2011). Proponents of local relativism for example, would limit themselves to question only certain areas such as aesthetics (e.g., an entity is beautiful or it is not beautiful). Other aspects such as scientific facts would then not be the focus of relativistic considerations, so that not all truths are understood as relative in an unrestricted way (Köbel 2011). The motivation behind Coeckelberghs relativistic understanding, is not to induce a debate about the particular features of an entity – although he is often understood in that way. Coeckelberghs relativism more so stems from a rejection of objectivism and ideas about a universal truth overall. Moral status is “relativistic” in his account in the sense that it emphasizes that every perception of an entity is the result of unique interpretation and personal perspective – rather than being relativistic in the sense of “anything goes”.

References

- Allhoff, F., Lin, P., Moor, J. H., & Weckert, J. (2007). *Nanoethics: the ethical and social implications of nanotechnology*. John Wiley & Sons.
- Ames, R. (1998). One-world natural cosmology. In *East Asian philosophy*. In *The Routledge Encyclopedia of Philosophy*. Taylor and Francis. Retrieved 18 Jun. 2023, from <https://www.rep.routledge.com/articles/overview/east-asian-philosophy/v-1/sections/one-world-natural-cosmology>. doi:10.4324/9780415249126-G218-1
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, 136(2), 151–173.
- Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge University Press.
- Andreotta, A. J. (2020). The hard problem of AI rights. *AI & Society*.
- Anscombe, G. E. M. (1957). *Intention*. Oxford: Blackwell
- Aristotle (1955). *The Ethics of Aristotle: The Nichomachaen Ethics*. (rev. ed.) (J. K. Thomson, trans.). New York: Viking. p. 104.
- Aristotle (1984). *Politics*. In J. Barnes (Ed.), *The Complete Works of Aristotle Vol. II*, (pp.1986–2129). Princeton: Princeton University Press.
- Asaad, R. R. (2021). Virtual reality and augmented reality technologies: A closer look. *International Research Journal of Science, Technology, Education, & Management (IRJSTEM)*, 1(2).
- Baker, P., & Potts, A. (2013). ‘Why do white people have thin lips?’ Google and the perpetuation of stereotypes via auto-complete search forms. *Critical discourse studies*, 10(2), 187-204.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209.
http://doi.org/http://dx.doi.org.proxy.lib.iastate.edu/10.1207/s15327957pspr0303_3
- Basl, J. (2014). Machines as moral patients we shouldn’t care about (yet): The interests and welfare of current machines. *Philosophy & Technology*, 27(1), 79–96.
- Basl, J., & Sandler, R. (2013). Three Puzzles Regarding the Moral Status of Synthetic Organisms. In G. E. Kaebnick & T.H. Murray (Eds.), *Synthetic Biology and Morality: Artificial Life and the Bounds of Nature* (pp. 89-106). Cambridge: MIT Press.

-
- Bazerman, M. H., Gino, F., Shu, L. L., & Tsay, C. J. (2011). Joint evaluation as a real-world tool for managing emotional assessments of morality. *Emotion review*, 3(3), 290-292.
- Benford, G., & Malartre, E. (2007). *Beyond Human: Living with Robots and Cyborgs*, New York: Forge Book.
- Bennett, B., & Daly, A. (2020). Recognising rights for robots: Can we? Will we? Should we? *Law, Innovation and Technology*, 12(1), 60–80. <https://doi.org/10.1080/17579961.2020.1727063>
- Bentham, J. (1789). *An introduction to the principles of Morals and legislation*.
- Benton, T. (1993). *Natural relations: Ecology, animal rights and social justice*. Verso.
- Bernstein, M. H. (1998). *On moral considerability: An essay on who morally matters*. Oxford University Press.
- Bertolini, A., & Aiello, G. (2018). Robot companions: A legal and ethical analysis. *The Information Society*, 34(3), 130-140.
- Bigelow, J. (1829). *Elements of Technology*, 2nd ed. Boston: Hilliard, Gray, Little & Wilkins.
- Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences*, 23(5), 365–368.
- Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2017). Like trainer, like bot? Inheritance of bias in algorithmic content moderation. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II 9* (pp. 405-415). Springer International Publishing.
- Biocca, F., & Delaney, B. (1995). Immersive virtual reality technology. *Communication in the age of virtual reality*, 15(32), 10-5555.
- Birch, T.H., (1993). Moral considerability and universal consideration. *Environmental Ethics* 15, 313–332.
- Birhane, A., & van Dijk, J. (2020). Robot rights? Let's talk about human welfare instead. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 207- 213).
- Blitz, M. (2014). Understanding Heidegger on technology. *The New Atlantis*, 63-80.
- Block, N. (1995). A confusion about a function of consciousness. *Behavioural and Brain Sciences*,

-
- Boyd, D. (2009). *Streams of Content, Limited Attention: The Flow of Information through Social Media*. "streams of content, limited attention: The flow of information through social media." <https://www.danah.org/papers/talks/Web2Expo.html>
- Brey, P., & Søraker, J. H. (2009). Philosophy of computing and information technology. In D. M. Gabbay, P. Thagard, J. Woods, & A. W. M. Meijers (Eds.), *Philosophy of technology and engineering sciences* (pp. 1341–1407). Oxford: Elsevier
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W W Norton & Co.
- Bryson, J. (2018). The Moral, Legal, and Economic Hazard of Anthropomorphizing Robots and AI. In *Robophilosophy/TRANSOR* (p. 11).
- Bryson, J. J. (2010). Robots Should Be Slaves. In Y. Wilks (Ed.), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues* (pp. 63–74). Amsterdam: John Benjamin. doi:10.1075/nlp.8.11bry
- Bunge, M. (1974). *Technology as applied science*. Springer Netherlands.
- Cafaro, P. (2004). *Thoreau's living ethics: Walden and the pursuit of virtue*. University of Georgia Press.
- Callicott, J. B. (1989). In *Defense of the Land Ethic: Essays in Environmental Philosophy*. Albany, NY: State University of New York Press.
- Calverley, D. J. (2011). Legal rights for machines: Some fundamental concepts. In M. Anderson & S. L. Anderson (Eds.), *Machine ethics* (pp. 213–227). Cambridge University Press.
- Cappuccio, M. L., Peeters, A., & McDonald, W. (2020). Sympathy for dolores: Moral consideration for robots based on virtue and recognition. *Philosophy & Technology*, 33(1), 9–31
- Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 41(1), 75–85.
- Cassell, E. J. (2004). *The Nature of Suffering and the Goals of Medicine*. 2nd ed. New York: Oxford University Press.
- Castelfranchi, C. (1995). Guaranties for Autonomy in Cognitive Agent Architecture. In M.J. Woolridge & N.R. Jennings (Eds.), *Intelligent Agents I*. Berlin: Springer.

-
- Cave, S., Nyrup, R., Vold, K., & Weller, A. (2019). Motivations and risks of machine ethics. *Proceedings of the IEEE*, 107(3), 562–574.
- Chalmers, D. J. (1997). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.
- Chalmers, D. J. (2014). Intuitions in philosophy: A minimal defense. *Philosophical Studies*, 171(3), 535-544.
- Chen, A. Y., Zhong, C., & Lu, T. K. (2015). Engineering living functional materials. *ACS synthetic biology*, 4(1), 8-11.
- Chinen, M. A. (2016). The co-evolution of autonomous machines and legal responsibility. *Virginia Journal of Law and Technology Association*, 20(2), 338–39
- Churchill, J. (2003). What Socrates said to Phaedrus: Reflections on technology and education. *The Midwest Quarterly*, 44(2), 211.
- Clark, A. (2001). Natural-born cyborgs?. In *Cognitive Technology: Instruments of Mind: 4th International Conference, CT 2001 Coventry, UK, August 6–9, 2001 Proceedings* (pp. 17-24). Berlin, Heidelberg: Springer.
- Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, 58(1), 7-19.
- Clark, A., 2014. *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.
- Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics Inf Technol* 12, 209–221. <https://doi.org/10.1007/s10676-010-9235-5>
- Coeckelbergh, M. (2011). You, robot: on the linguistic construction of artificial others. *AI & society*, 26, 61-69.
- Coeckelbergh, M. (2012). *Growing Moral Relations: Critique of Moral Status Ascription*. London: Palgrave. doi:10.1057/9781137025968
- Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-cartesian moral hermeneutics. *Philosophy & Technology*, 27(1), 61–77. <https://doi.org/10.1007/s13347-013-0133-8>
- Coeckelbergh, M. (2018). Why care about robots? Empathy, moral standing, and the language of suffering. *Kairos Journal of Philosophy & Science*, 20(1), 141–158.

-
- Coeckelbergh, M. (2021). How to Use Virtue Ethics for Thinking About the Moral Standing of Social Robots: A Relational Interpretation in Terms of Practices, Habits, and Performance. *Int J of Soc Robotics* 13, 31–40.
- Cointe, N., Bonnet, G., & Boissier, O. (2016). Ethical Judgment of Agents' Behaviors in Multi-Agent Systems. In AAMAS (pp. 1106-1114).
- Collingwood, R. G. (1960). *The idea of nature*. Oxford University Press.
- Cooley, D. R., & Cooley, D. R. (2010). Are Transgenic Organisms, Biotechnology, and Technology Unnatural?. *Technology, Transgenics and a Practical Moral Code*, 93-140
- Copp, D. (2005). *The Oxford handbook of ethical theory*. Oxford University Press.
- Coulton, P., & Lindley, J. G. (2019). More-than human centred design: Considering other things. *The Design Journal*, 22(4), 463-481.
- Croes, E. A., & Antheunis, M. L. (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1), 279-300.
- Crystal, D. (2008). Texting. *Elt journal*, 62(1), 77-83.
- Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*, 26(4), 2023–2049.
<https://doi.org/10.1007/s11948-019-00119-x>
- Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. *Robot Law, Calo, Froomkin, Kerr eds., Edward Elgar*.
- Davidson, D. (1963). Actions, reasons, and causes. *The journal of philosophy*, 60(23), 685-700.
- Davies, C. R. (2011). An evolutionary step in intellectual property rights—Artificial intelligence and intellectual property. *Computer Law & Security Review*, 27(6), 601–619.
- De Vries, G. (1999). *Zeppelins Over Filosofie, Technologie En Cultuur*.
- De Vries, G. H. (1999). *Zeppelins-Over filosofie, technolgie en cultuur*. AmsterdamVan Gennep.
- DeGrazia, D. (2016), Modal Personhood and Moral Status: A Reply to Kagan's Proposal. *J Appl Philos*, 33: 22-25. <https://doi.org/10.1111/japp.12166>

-
- Dennett, D. C. (1996). Facing backwards on the problem of consciousness. *Journal of Consciousness Studies*, 3, 4-6.
- Dennett, D. C. (1998). *Brainchildren: Essays on designing minds*. MIT Press.
- Descartes, R. (1637). *Discours de la méthode pour bien conduire sa raison et chercher la verité dans les sciences*. Leiden
- Dixon, E. (2015). *Constructing the identity of AI: A discussion of the AI debate and its shaping by science fiction*. Leiden University.
- Donovan, J., & Adams, C. J. (1996). Beyond animal rights: A feminist caring ethic for the treatment of animals.
- Dracopoulou, S. (2003). The ethics of creating conscious robots—Life, personhood and bioengineering. *Journal of Health, Social and Environmental Issues*, 4(2), 47–50.
- Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human factors*, 40(3), 354-375.
- Draude, C. (2020). “Boundaries Do Not Sit Still” from Interaction to Agential Intra-action in HCI. In *Human-Computer Interaction. Design and User Experience: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22* (pp. 20-32). Springer International Publishing.
- Dreyfus, H. L. (1997). Heidegger on gaining a free relation to technology. *Technology and values*, 4, 1-54.
- Dreyfus, H. L., (1990), *Being-in-the-World: A Commentary on Heidegger's Being and Time*, Division I, Cambridge, Mass.: MIT Press.
- Drummond, J. (2015). Intentionality, phenomenological perspectives. In *The Routledge Encyclopedia of Philosophy*. Taylor and Francis. Retrieved 13 May. 2023, from <https://www.rep.routledge.com/articles/thematic/intentionality-phenomenological-perspectives/v-1>
- Ellul, J. (1964). *The technological society*. Vintage Books.
- Epstein, R., Mohr Jr, R., & Martinez, J. (2018). The Search Suggestion Effect (SSE): How Search Suggestions Can Be Used to Shift Opinions and Voting Preferences Dramatically and Without People’s Awareness’. In *98th Annual Meeting of the Western Psychological Association*.

-
- Esposito, E. (2022). Artificial Communication? Algorithms as Interaction Partners. *Artificial Communication*.
- Ess, C. (2015). The onlife manifesto: Philosophical backgrounds, media usages, and the futures of democracy and equality. In L. Floridi, *The onlife manifesto: Being human in a hyperconnected era*, (pp. 89-109).
- Estrada, D. (2018). Value alignment, fair play, and the rights of service robots. In *Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society* (pp. 102–107). Presented at the AIES '18: AAAI/ ACM conference on AI, ethics, and Society. ACM.
<https://doi.org/10.1145/3278721.3278730>
- Estrada, D. (2020). Human supremacy as posthuman risk. *Journal of Sociotechnical Critique*, 1(1), 1–40. <https://doi.org/10.25779/J5PS-DY87>
- Feenberg, A. (1991). *Critical Theory of Technology*. Oxford: Oxford University Press.
- Feenberg, A. (2003). Active and passive bodies: Comments on Don Ihde's *Bodies in Technology*. *Techné: Research in Philosophy and Technology*, 7(2), 125-130.
- Feenberg, A. (2010). *Between reason and experience: Essays in technology and modernity*. MIT Press.
- Ferré, F. (1988). *Philosophy of Technology*, Englewood Cliffs (NJ): Prentice Hall.
- Floridi, L. (1999). Information ethics: On the philosophical foundations of computer ethics. *Ethics and Information Technology*, 1(1), 33–52. <https://doi.org/10.1023/A:1010018611096>
- Floridi, L. (2004). Open problems in the philosophy of information. *Metaphilosophy* 35(4), 554–582.
- Floridi, L. (2007). A look into the future impact of ICT on our lives. *The Information Society* 23(1), 59–64.
- Floridi, L. (2008). Foundations of Information Ethics. In K. E. Himma & H. T. Tavani (Eds.), *The Handbook of Information and Computer Ethics* (pp. 1-23). Wiley.
- Floridi, L. (2013). Distributed morality in an information society. *Science and engineering ethics*, 19, 727-743.
- Floridi, L. (2015). *The onlife manifesto: Being human in a hyperconnected era*. Springer Nature.

-
- Floridi, L., Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines* 14, 349–379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>
- Frederik, I., Sonneveld, W. & de Vries, M.J. (2011). Teaching and learning the nature of technical artifacts. *Int J Technol Des Educ* 21, 277–290. <https://doi.org/10.1007/s10798-010-9119-3>
- Friedman, C. (2020). Human-robot moral relations: Human interactants as moral patients of their own agential moral actions towards robots. In *Artificial Intelligence Research: First Southern African Conference for AI Research, SACAIR 2020, Muldersdrift, South Africa, February 22-26, 2021, Proceedings 1* (pp. 3-20). Springer International Publishing.
- Ganascia, J. (2015). Views and Examples on Hyper-Connectivity. 10.1007/978-3-319-04093-6_13.
- Garber, M. (2013). Funerals for Fallen Robots. [Online]. The Atlantic. Last updated: 20 September 2013. Available at: <https://www.theatlantic.com/technology/archive/2013/09/funerals-for-fallen-robots/279861/> [Accessed 16 December 2021].
- Gerdes, A. (2016). The issue of moral consideration in robot ethics. *Acm Sigcas Computers and Society*, 45(3), 274-279.
- Gibson, W. (2019). *Neuromancer* (1984). In *Crime and Media* (pp. 86-94). Routledge.
- Gier, N. (1980). Wittgenstein and Forms of Life. *Philosophy of the Social Sciences*, 10(3), 241–258.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge: Harvard University Press.
- Giroux V. & Larue, R. (2015) Pathocentrisme. In D. Bourg & A. Papaux (Eds.), *Dictionnaire de la pensée écologique*. Presses universitaires de France, Paris
- Goertzel, B. (2002). Thoughts on AI morality. *Dynamical Psychology: An International, Interdisciplinary Journal of Complex Mental Processes* (May).
- Goodpaster, K. E. (1978). On Being Morally Considerable. *The Journal of Philosophy*, 75(6), 308–25. <https://doi.org/10.2307/2025709>
- Gordon, J. S. (2020). What do we owe to intelligent robots?. In *Smart Technologies and Fundamental Rights* (pp. 17-47). Brill.
- Gordon, J. S., & Gunkel, D. J. (2022). Moral status and intelligent robots. *The Southern Journal of Philosophy*, 60(1), 88-117.

-
- Gowans, C. (2021), Moral Relativism, In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). <https://plato.stanford.edu/archives/spr2021/entries/moral-relativism/>
- Gray, K., Young, L. & Waytz, A. (2012) Mind Perception Is the Essence of Morality. *Psychological Inquiry*, 23(2), 101-124. DOI: [10.1080/1047840X.2012.651387](https://doi.org/10.1080/1047840X.2012.651387)
- Greitemeyer, T., & Mügge, D. O. (2014). Video games do affect social outcomes: A meta-analytic review of the effects of violent and prosocial video game play. *Personality and Social Psychology Bulletin*, 40(5), 578–589. <http://doi.org/10.1177/0146167213520459>
- Gruen, L. (2003). The moral status of animals. *Stanford encyclopaedia of philosophy*. Retrieved from <http://plato.stanford.edu/entries/moral-animal/>.
- Gunkel, D. J. (2007). Thinking otherwise: Ethics, technology and other subjects. *Ethics and Information Technology*, 9(3), 165–177. <https://doi.org/10.1007/s10676-007-9137-3>
- Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. The MIT Press. <https://doi.org/10.7551/mitpress/8975.001.0001>
- Gunkel, D. J. (2022). The relational turn: Thinking robots otherwise. *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds With Robots*. Edited by Janina Loh and Wulf Loh. Transcript Verlag. Forthcoming.
- Gunkel, D.J. (2018). The other question: can and should robots have rights?. *Ethics Inf Technol* 20, 87–99. <https://doi.org/10.1007/s10676-017-9442-4>
- Gutman, M., Rathgeber, B., & Syed, T. (2012). Action and autonomy: A hidden dilemma in artificial autonomous systems. *Robo-and Informationethics. Some Fundamentals*, 231-256.
- Habermas, J. (1984). *The Theory of Communicative Action*. Boston: Beacon.
- Habermas, J. (2019). *Auch eine Geschichte der Philosophie*, Bd. 1. 1. Auflage. Berlin: Suhrkamp.
- Hacker-Wright, J. (2007). Moral Status in Virtue Ethics. *Philosophy*, 82(3), 449-473.
- Haila, Y. (2000). Beyond the nature-culture dualism. *Biology and philosophy*, 15, 155-175.
- Hajdin, M. (1994). *The boundaries of moral discourse*.
- Hale, B. (2011). Moral Considerability: Deontological, not Metaphysical. *Ethics and the Environment*, 16(2), 37–62. <https://doi.org/10.2979/ethicsenviro.16.2.37>
- Hall, J. S. (2007). *Beyond AI: Creating the conscience of the machine*. Prometheus books.

-
- Haraway, D. (1991). A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century. In D. Haraway (Ed.), *Simians, Cyborgs and Women: The Reinvention of Nature*, (pp. 149–81). New York: Routledge.
- Hare, R. M.. (1996). Foundationalism and Coherentism in Ethics, Moral Knowledge? In W. Sinnott-Armstrong and M. Timmons (Eds.), *New Readings in Moral Epistemology* (pp. 190-199). Oxford: Oxford University Press
- Harman, G. (1986). Moral Explanations of Natural Facts—Can Moral Claims be Tested Against Moral Reality?. *Southern Journal of Philosophy* 24, 57-68.
- Harris, J., Anthis, J. R. (2021). The Moral Consideration of Artificial Entities: A Literature Review. *Science and Engineering Ethics*, 27-53. <https://doi.org/10.1007/s11948-021-00331-8>
- Haugeland, J. (1981). *Mind Design*. Montgomery, VT: Bradford Books.
- Heidegger, M. (1993). *The question concerning technology*.
- Heidegger, M. (1996). *Being and time: A translation of Sein und Zeit*. SUNY press.
- Hiller, A. (2011). Climate change and individual responsibility. *The Monist*, 94(3), 349-368.
- Himma, K. E. (2003). The relationship between the uniqueness of computer ethics and its independence as a discipline in applied ethics. *Ethics and Information Technology*, 5(4), 225–237.
- Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent?. *Ethics and Information Technology*, 11, 19-29.
- Höök, K. (2009). Affective loop experiences: designing for interactional embodiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3585-3595.
- Hoły-Łuczaj, M., & Blok, V. (2019). How to deal with hybrids in the Anthropocene? Towards a philosophy of technology and environmental philosophy 2.0. *Environmental Values*, 28(3), 325-345.
- Horgan, T., & Timmons, M. (2008). WHAT DOES MORAL PHENOMENOLOGY TELL US ABOUT MORAL OBJECTIVITY? *Social Philosophy and Policy*, 25(1), 267-300. doi:10.1017/S0265052508080102

-
- Hursthouse, R. (2007). Environmental virtue ethics. *Working virtue: Virtue ethics and contemporary moral problems*, 155-171.
- Husserl, E. (1913). *Ideas: General Introduction to Pure Phenomenology* (W. R. B. Gibson, trans.). New York: Collier Books
- Idhe, D. (1990). *Technology and the lifeworld: From garden to earth*. Indianapolis, IN: Indiana University Press.
- Idhe, D. (1993). *Postphenomenology*. Evanston, IL: Northwestern University Press.
- Ingold, T. (1996). Human worlds are culturally constructed: against the motion (I). In T. Ingold (Ed.), *Key debates in anthropology* (pp. 112-118). London: Routledge
- Introna, L. D. (2007, May). Towards a post-human intra-actional account of socio-technical agency (and morality). In *Proceedings of the Moral agency and technical artefacts scientific workshop (NIAS, Hague, 10-12 May 2007, 2007)*
- Jaworska, A., & Tannenbaum, J. (2013). The grounds of moral status.
- Jaynes, T. L. (2021). The question of algorithmic personhood and being (or: on the tenuous nature of human status and humanity tests in virtual spaces—why all souls are ‘necessarily’ equal when considered as energy). *J*, 4(3), 452-475.
- Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology* 8, 195–204.
- Johnson, D. G., & Verdicchio, M. (2018). Why robots should not be treated like animals. *Ethics and Information Technology*, 20(4), 291–301.
- Johnson, D.G. (1985). *Computer Ethics*. Upper Saddle River, NJ: Prentice Hall.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.
- Kaiser, C., Ahuvia, A., Rauschnabel, P. A. & Wimble, M. (2020). Social media monitoring: What can marketers learn from Facebook brand photos?. *Journal of Business Research*, 117, 707–717.
- Kamm, F. M. (2008). *Intricate ethics: rights, responsibilities, and permissible harm*. Oxford University Press.
- Kant, I. (1785). *Groundwork of the metaphysics of morals*.

-
- Kant, I. (1788). *Kritik der reinen Vernunft*.
- Kant, I. (1963). Duties to animals and spirits. *Lectures on ethics*, 212, 458-459.
- Kapor, M., & Barlow, J. P. (1990). *Across the electronic frontier*. Electronic Frontier Foundation, 10.
- Kölbel, M. (2011). Global Relativism and Self-Refutation. *A companion to relativism*, 9-30.
- Korsgaard, Christine, 1983. "Two Distinctions in Goodness", *The Philosophical Review* 92(2): 169-95.
- Kurzweil, R. (2005). The singularity is near. In *Ethics and emerging technologies* (pp. 393-406). London: Palgrave Macmillan UK.
- Kwok, C., Grisham, J. R., & Norberg, M. M. (2018). Object attachment: Humanness increases sentimental and instrumental values. *Journal of behavioral addictions*, 7(4), 1132-1142.
- Latour, B. (1993). *We Have Never Been Modern*, trans. Catherine Porter. Cambridge, MA: Harvard University Press.
- Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale Welt*, 369-381.
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Latour, B., & Venn, C. (2002). Morality and technology. *Theory, culture & society*, 19(5-6), 247-260.
- Laukyte, M. (2017). Artificial agents among us: Should we recognize them as agents proper? *Ethics and Information Technology*, 19(1), 1–17. <https://doi.org/10.1007/s10676-016-9411-3>
- Laukyte, M. (2019). AI as a Legal Person. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (pp. 209-213).
- Lender, L. (2016). Weighing the moral interests of AI.
- Leopold, A. (1966). *A Sand County Almanac with Essays on Conservation from Round River*. New York: Random House.
- Levinas, E. (1979). *Totality and infinity: An essay on exteriority*(Vol. 1). Springer Science & Business Media.
- Levy, D. (2009). The ethical treatment of artificially conscious robots. *International Journal of Social Robotics*, 1(3), 209–216. <https://doi.org/10.1007/s12369-009-0022-6>

-
- Li, C. (2014). The Confucian conception of freedom. *Philosophy East and West*, 902-919.
- Lima, G., Zhunis, A., Manovich, L., & Cha, M. (2021). On the social-relational moral standing of AI: An Empirical study using AI-generated art. *Frontiers in Robotics and AI*, 8, 719944.
- Lin, P., Abney, K., & Bekey, G. (2011). Robot ethics: Mapping the issues for a mechanized world. *Artificial intelligence*, 175(5-6), 942-949
- Loughnan, S., Haslam, N., & Bastian, B. (2010). The role of meat consumption in the denial of moral status and mind to meat animals. *Appetite*, 55(1), 156–159.
<https://doi.org/10.1016/j.appet.2010.05.043>
- Lyotard, J. 1984. *The Postmodern Condition: A Report on Knowledge*. Trans. Geoff Bennington & Brian Massumi. Minneapolis, MN: University of Minnesota Press.
- MacIntyre, A. (1984). The claims of after virtue. *Analyse & Kritik*, 6(1), 3-7.
- Mackenzie, R. (2014). Sexbots: replacements for sex workers? Ethical constraints on the design of sentient beings for utilitarian purposes. In *Proceedings of the 2014 workshops on advances in computer entertainment conference-ACE '14 workshops* (pp. 1–8). Presented at the 2014 workshops. ACM Press. <https://doi.org/10.1145/2693787.2693789>
- Malle, B. F. (2016). Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics and Information Technology*, 18(4), 243–256.
- Marinov, R. (2019) New Intelligent Technologies—Interactivity and Information Issues. *Advances in Journalism and Communication*, 7, 94-108.
- McCarthy J, Minsky ML, Rochester N, Shannon CE (2006) A proposal for the Dartmouth summer research project on artificial intelligence. *AI Mag* 27:12–14. <https://doi.org/10.1609/aimag.v27i4.1904>
- McLuhan, M. (1994). *Understanding media: The extensions of man*. Cambridge, MA: MIT Press.
- McLuhan, M., Powers, B. R., Leonhardt, C. P., & Baacke, D. (1995). *The global village: der Weg der Mediengesellschaft in das 21. Jahrhundert*. Junfermann.
- McMahan, J. (2005). Our Fellow Creatures. *The Journal of Ethics* 9, 353–380.
- McPherson, T. (1984). The moral patient. *Philosophy* 59(228), 171–183.

-
- Merleau-Ponty, M., & Smith, C. (1962). *Phenomenology of perception* (Vol. 2012). London: Routledge.
- Metz, C. (2011, February 18). *Google opens curtain on "Manual" search penalties*. The Register® - Biting the hand that feeds IT.
https://www.theregister.com/2011/02/18/google_on_manual_search_penalties/
- Miller, H. B. (1994). Science, ethics, and moral status. *Between the Species: A Journal of Ethics* 10(1), 10–18.
- Minsky, M. (1980). *Telepresence*.
- Mitcham, C. (1978). Types of technology. *Research in philosophy and technology*, 1(1), 229-294.
- Mitcham, C. (1994): *Thinking Through Technology: The Path Between Engineering and Philosophy*. Chicago & London: University of Chicago Press.
- Molina, M. (2022). What is an intelligent system?.
- Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, 16(4), 266–275.
- Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems* 21(4), 18–21.
- Mosakas, K. (2020). On the moral status of social robots: Considering the consciousness criterion. *AI & Society*, 36, 429-443. <https://doi.org/10.1007/s00146-020-01002-1>
- Moser, P. K. & Carson, T. L. (2001). *Moral relativism; a reader*. Oxford University Press
- Müller, V.C. (2021). Is it time for robot rights? Moral status in artificial entities. *Ethics Inf Technol* 23, 579–587. <https://doi.org/10.1007/s10676-021-09596-w>
- Naess, A., (1995). Self-realization. An ecological approach to being in the world, In G. Sessions (Ed.), *Deep Ecology for the Twenty-First Century* (pp. 225-239). Boston and London: Shambhala
- Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, 83(4), 435-450.
- Nagel, T. (1989). *The view from nowhere*. Oxford university press.
- Natanson, M. A. (1970). *The journeying self: A study in philosophy and social role*.

-
- Neely, E. L. (2014). Machines and the moral community. *Philosophy & Technology*, 27(1), 97–111. <https://doi.org/10.1007/s13347-013-0114-y>
- Nick, E. (2021, October 6). *How artificial intelligence is powering search engines*. Data Science Central. <https://www.datasciencecentral.com/how-artificial-intelligence-is-powering-search-engines/>
- Nietzsche, F. (1966). *Beyond good and evil* (W. Kaufmann, Trans.). *New York: Vintage, 1989*.
- Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, USA: New York University Press. <https://doi.org/10.18574/nyu/9781479833641.001.0001>
- Nye, H., & Yoldas, T. (2020). Artificial moral patients: mentality, intentionality, and systematicity. *The International Review of Information Ethics*, 29.
- Nyholm, S. (2023). Tools and/or Agents? Reflections on Sedlakova and Trachsel's Discussion of Conversational Artificial Intelligence. *The American Journal of Bioethics*, 23(5), 17-19.
- O'Neil, R. (1997). Intrinsic Value, Moral Standing, and Species. *Environmental Ethics* 19(1), 45-52.
- O'Regan, J. K. (2007). How to build consciousness into a robot: the sensorimotor approach. *50 Years of Artificial Intelligence: Essays Dedicated to the 50th Anniversary of Artificial Intelligence*
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK. nce, 332-346.
- Rabinow, P. (1992). Studies in the Anthropology of Reason. *Anthropology Today*, 8(5), 7-10.
- Rapp, F. (1989). Introduction: General perspectives on the complexity of philosophy of technology. In P. T. Durbin (Ed.), *Philosophy of Technology: Practical, Historical and Other Dimensions* (pp. iX-XXI). Dordrecht: Kluwer
- Reeve, C. D. (2004). *Plato: Republic*. Hackett, Indianapolis.
- Regan, T. (1983). *The case for animal rights*. Berkeley: The University of California Press.
- Reydon, T. A. C. (n. d.), Philosophy of Technology. In J. Fieser, J. & B. Dowden, *Internet Encyclopedia of Philosophy*. [Philosophy of Technology | Internet Encyclopedia of Philosophy \(utm.edu\)](https://www.iep.utm.edu/philosophy-technology/) (retrieved on: 14.5.2023).
- Richardson, K. (2016). Sex robot matters: Slavery, the prostituted, and the rights of machines. *IEEE Technology and Society Magazine*, 35(2), 46–53. <https://doi.org/10.1109/MTS.2016.255442>

-
- Robertson, J. (2014). Human rights versus robot rights: Forecasts from Japan. *Critical Asian Studies*, 46(4), 571–598. <https://doi.org/10.1080/14672715.2014.960707>
- Ropohl, G. (1990). Technisches Problemlösen und soziales Umfeld. In F. Rapp (Ed.), *Technik und Philosophie* (pp. 111-167). Düsseldorf: VDI.
- Rosemont, H., & Ames, R. T. (1998). *The analects of confucius: A philosophical translation*. New York: Ballantine.
- Rouillard, J. (2012). The Pervasive Fridge. A smart computer system against uneaten food loss. In *Seventh international conference on systems (ICONS2012)* (pp. pp-135).
- Sætra, H. S. (2021). Challenging the neo-anthropocentric relational approach to robot rights. *Frontiers in Robotics and AI*, 8, 744426.
- Sætra, H. S. & Danaher, J. (2022). To Each Technology Its Own Ethics: The Problem of Ethical Proliferation. *Philos. Technol.* 35, 93. <https://doi.org/10.1007/s13347-022-00591-7>
- Sætra, H. S., & Fosch-Villaronga, E. (2021). Research in AI has Implications for Society: How do we Respond? *Morals & Machines*, 1(1), 60-73.
- Samuelsson, L. (2008). *The moral status of nature : reasons to care for the natural world* (PhD dissertation, Institutionen för idéoch samhällsstudier, Umeå universitet). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-1612>
- Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, 22(1), 32–70. <https://doi.org/10.1177/1088868317698288>
- Schmetkamp, S. (2020). Understanding A.I.—Can and should we empathize with robots? *Review of Philosophy and Psychology*, 11(4), 881–897. <https://doi.org/10.1007/s13164-020-00473-x>
- Searle, J. R., Dennett, D. C., & Chalmers, D. J. (1997). *The mystery of consciousness*. New York: Review of Books.
- Searle, J., (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, 417–57
- Shen, H., DeVos, A., Eslami, M., & Holstein, K. (2021). Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-29.
- Shepherd, J. (2018). *Consciousness and Moral Status*. Routledge

-
- Simondon, G. (1989). *Du mode d'existence des objets techniques*.
- Singer, P. (1975). *Animal Liberation: A new ethics for our treatment of animals*. Harper Collins
- Singer, P. (2003). Voluntary Euthanasia: A Utilitarian Perspective. *Bioethics*, 17, 526-541.
- Sinnott-Armstrong, W. (2005). It's Not My. In *Perspectives on climate change: Science, economics, politics, ethics* (Vol. 5, pp. 285-307). Emerald Group Publishing Limited.
- Siponen, M. (2004). A pragmatic evaluation of the theory of information ethics. *Ethics and Information Technology*, 6, 279-290.
- Skrbina, D. (2014). *The metaphysics of technology*. Routledge.
- Smith, David Woodruff, "Phenomenology", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.)
- Søraker, J. H. (2007). The Moral Status of Information and Information Technologies.
- Souriau, Etienne (1943) *Les différents modes d'existence*. Paris: PUF.
- Sparrow, R. (2004). The turing triage test. *Ethics and Information Technology*, 6(4), 203–213.
- Sparrow, R. (2020). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics*.
- St Ledger, U., Reid, J., Begley, A., Dodek, P., McAuley, D. F., Prior, L., & Blackwood, B. (2021). Moral distress in end-of-life decisions: a qualitative study of intensive care physicians. *Journal of critical care*, 62, 185-189.
- Steiner, G. (2010). *Anthropocentrism and its Discontents: The Moral Status of Animals in the History of Western Philosophy*. University of Pittsburgh Press: Pittsburgh, PA, USA.
- Stevenson, C. L. (1994). *Ethics and Language*. New Haven: Yale University Press.
- Stone, C. D. (1972). Should trees have standing--toward legal rights for natural objects. *S. CAL. L. rev.*, 45, 450.
- Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press.
- Sullins, J. P. (2005). Ethics and Artificial life: From Modeling to Moral Agents. *Ethics Inf Technol* 7, 139–148. <https://doi.org/10.1007/s10676-006-0003-5>

-
- Summers, C. (2016). Can ‘Samantha’ vote? On the question of singularity, citizenship and the franchise. Presented at the humanities and technology association conference.
- Sutherland, N. S. (1996). *The international dictionary of psychology*. Crossroad Publishing Company.
- Swoyer, C., 2010, “Relativism”, *The Stanford Encyclopedia of Philosophy*, (Winter 2010 Edition), Edward N. Zalta (ed.).
- Tannenbaum, J. & Jaworska, A. (2018). The Grounds of Moral Status. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition).
- Tavani, H. (2018). Can social robots qualify for moral consideration? Reframing the question about robot rights. *Information*, 9(4), 73. <https://doi.org/10.3390/info9040073>
- Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Cambridge, MA: Harvard University Press.
- Taylor, P. (1986). *Respect for nature: a theory of environmental ethics*. Princeton: Princeton University Press.
- Thomas, O. (2018). Two decades of cognitive bias research in entrepreneurship: what do we know and where do we go from here?. *Management Review Quarterly*, 68(2), 107-143.
- Tigard, D.W. (2021). Responsible AI and moral responsibility: a common appreciation. *AI Ethics* 1, 113–117. <https://doi.org/10.1007/s43681-020-00009-0>
- Tollon, F. (2020). The artificial view: Toward a non-anthropocentric account of moral patiency. *Ethics and Information Technology*. <https://doi.org/10.1007/s10676-020-09540-4>
- Tollon, F. & Naidoo, K. (2021). On and beyond artifacts in moral relations: accounting for power and violence in Coeckelbergh’s social relationism. *AI & Soc.*
- Tomasik, B. (2014). Do artificial reinforcement-learning agents matter morally? Center on Long-Term Risk. <https://longtermrisk.org/do-artificial-reinforcement-learning-agents-matter-morally/>
- Torrance, S. (2008). Ethics and consciousness in artificial agents. *AI & Society*, 22(4), 495–521. <https://doi.org/10.1007/s00146-007-0091-8>
- Torrance, S. (2013). Artificial agents and the expanding ethical circle. *AI & Society*, 28(4), 399–414. <https://doi.org/10.1007/s00146-012-0422-2>

-
- Tzafestas, S. G. (2016). Roboethics: A branch of applied ethics. In S. G. Tzafestas (Ed.), *Roboethics: A navigating overview* (pp. 65–79). Springer. https://doi.org/10.1007/978-3-319-21714-7_5
- Umbrello, S. & Sorgner, S. L. (2019). Nonconscious cognitive suffering: Considering suffering risks of embodied artificial intelligence. *Philosophies*, 4(2), 24. <https://doi.org/10.3390/philosophies4020024>
- Vanderheiden, Steve. 2007. "Climate Change and the Challenge of Moral Responsibility." *Journal of Philosophical Research* 32: 85–92
- Velmans, M. (2000). *Understanding Consciousness*. New York: Routledge.
- Venturini, F. (2022). Intelligent technologies and productivity spillovers: Evidence from the Fourth Industrial Revolution. *Journal of Economic Behavior & Organization*, Elsevier, 194(C), 220-243.
- Verbeek, P. P. (2005). *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, Pennsylvania: Penn State University Press.
- Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things*. University of Chicago press.
- Verbeek, P. P. (2015). Designing the public sphere: information technologies and the politics of mediation. In L. Floridi, *The Onlife manifesto: being human in a hyperconnected era* (pp. 217-227).
- Verrugio, P.M. & Operto, F. (2006). Roboethics: A bottom -up interdisciplinary discourse in the field of applied ethics in robotics. *IRIE Int. Rev. Inf. Ethics*, 6, 2–8.
- Wallach, W., Allen, C., & Smith, I. (2008). Machine morality: Bottom-up and top-down approaches for modelling human moral faculties. *AI & Society*, 22(4), 565–582.
- Warnock, G. J. (1971). *The object of morality*. Methuen.
- Warren, M. A. (1997). *Moral status: Obligations to persons and other living things*. Clarendon Press.
- Wassmann, J., Träuble, B., & Funke, J. (2013). Theory of mind in the Pacific: Reasoning across cultures.
- Weiser, M. (1991). The Computer for the 21st Century. *Scientific American*, 265(3), 94-105.
- Weiss, G. (1999). *Multiagent Systems*. MIT Press.

-
- Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, 25(1), 166.
- Wetlesen, J. (1999). The moral status of beings who are not persons: A casuistic argument. *Environmental Values*, 8, 287-323.
- Wheeler, M. (2011). Martin Heidegger. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition). <http://plato.stanford.edu/archives/win2011/entries/heidegger/>
- Winch, P. (1958). *The Idea of a Social Science and its Relation to Philosophy*, 2nd edition. London: Routledge, 1990.
- Winner, L. (1977). *Autonomous Technology: Technics-out-of-control as a Theme in Political Thought*. Cambridge (MA): MIT Press.
- Wittgenstein, L. (1953). *Philosophical Investigations*, 4th edition (trans. G.E.M. Anscombe, M.S. Hacker & J. Schulte). Oxford: Wiley-Blackwell, 2009.
- Wolfendale, J. (2007). My avatar, my self: Virtual harm and attachment. *Ethics and information technology*, 9, 111-119.
- Wooldridge, M., & Jennings, N. (1994). Agent Theories, Architectures, and Languages: A Survey. In *Proceedings of the International Workshop on Agent Theories, Architectures & Languages (ECAI'94)*. Springer.
- Wu, T. (2012). Machine speech. *University of Pennsylvania Law Review*, 161, 1495–1533.
- Zhong, N. (2001). *Intelligent agent technology: Research and development*. World Scientific.
- Złotowski, J., Proudfoot, D., Yogeewaran, K., & Bartneck, C. (2015). Anthropomorphism: opportunities and challenges in human–robot interaction. *International journal of social robotics*, 7, 347-36