Chapter 12
Artificial Intelligence and African Conceptions of Personhood

C. S. Wareham

Abstract Under what circumstances if ever ought we to grant that artificial intelligence (AI) is a person? The question of whether AI could have the high degree of moral status that is attributed to human persons has received little attention. What little work there is employs Western conceptions of personhood, while non-Western approaches are neglected. In this chapter, I discuss African conceptions of personhood and their implications for the possibility of AI persons. I focus on an African account of personhood that is prima facie inimical to the idea that AI could ever be 'persons' in the sense typically attributed to humans. I argue that despite its apparent anthropocentrism, this African account could admit AI as persons.

Keywords Artificial intelligence · Moral status · Personhood · African ethics · Anthropocentrism

12.1 Introduction

Machine learning and computational intelligence perform increasingly significant social roles. Unsurprisingly then, there is a growing literature regarding their moral status, with theorists such as Floridi suggesting it is justified to regard artificial agents as having intrinsic moral value for their own sake (Floridi & Sanders, 2004). While issues about intrinsic value are important, the question of whether artificial

This chapter is a reprint of "Wareham, C.S. Artificial intelligence and African conceptions of personhood. *Ethics Inf Technol* **23**, 127–136 (2021)".

C. S. Wareham (✉)
The Ethics Institute, Department of Philosophy and Religion, Utrecht University, Utrecht, the Netherlands
e-mail: c.s.wareham@uu.nl

A. D. Attoe et al. (eds.), *Conversations on African Philosophy of Mind, Consciousness and Artificial Intelligence*,
https://doi.org/10.1007/978-3-031-36163-0_12

intelligence (AI) could have the high degree of moral status that is attributed to human persons has received little attention. Moreover, what little work there is employs Western conceptions of personhood (Coeckelbergh, 2010a; Wareham, 2011), while non-Western approaches are neglected. In this chapter, I examine an African account of personhood that is prima facie inimical to the idea that AI could ever be 'persons' in the sense typically attributed to humans. I argue that despite its apparent anthropocentrism, this African account could allow for AI persons.

In making this claim, I should point out three limitations at the outset. The first is that I will not present a strong case for the claim that there are or could ever in fact be artificial agents capable of duplicating human cognitive behaviour. While I present some reasons to think this might occur, the question of whether such beings could actually exist has generated enormous debate that it is impossible to engage with here. The aim is instead to consider the circumstances under which, if we were presented with AI agents, we should, on the basis of an African conception of personhood, consider them as persons with all the relevant rights and duties that this entails. The second limitation regards the African account of personhood. I do not claim that this is the only African account of personhood, or that it is superior to Western accounts. While I will mention some potential criticisms and strengths, my aim is to apply the account, rather than to critique or defend it. The third limitation concerns the implications of moral personhood for legal personhood. The relation between these is complex. Though arguably the latter should follow the former, I will not make this claim nor discuss these implications in any detail as to do so would go beyond my current scope.

I begin describing some avenues of research in AI, before homing in on the conception of personhood with which this chapter will be concerned as 'threshold personhood'. Thereafter, I suggest that the non-anthropocentric nature of Western threshold accounts could in principle permit AI. By contrast, I point out that African accounts of personhood are typically anthropocentric. I outline perhaps the most comprehensive African-inspired account of moral status, according to which attribution of highest moral status to humans stems from capacities for mutual recognition as both objects and subjects of harmonious relationships (Metz, 2012). Prima facie, this account presents special difficulties for potential personhood of AI due to anthropocentric elements of African accounts. However, I claim that empirical evidence suggests that these difficulties can be overcome. In principle, AI could be regarded as persons with equivalent moral status. I conclude by discussing the implications of this.

12.2 Artificial Intelligence Research

Uses of robots have diversified to include warfare, education, entertainment, sex, and healthcare (Coeckelbergh, 2010b). Inevitably their increased social importance has generated interest in potential applications of AI. This in turn has generated numerous ethical questions concerning justified and unjustified uses and the

potential dangers presented by AI. Interest has focused on the types of moral rules artificial agents should have (Allen et al., 2000; Etzioni & Etzioni, 2017) and how these rules could be acquired (Allen et al., 2005). Theorists have also discussed the moral status of artificial agents, that is, whether AI should be treated as objects of moral concern (Brey, 2008; Versenyi, 1974). There are also significant concerns about the responsibility of and for artificial agents (Floridi & Sanders, 2004).

However, despite the increasing interest in moral issues surrounding AI, few theorists have considered whether artificial agents could achieve equivalent moral status to that of human persons. This gap may be because the relevant sort AI has hitherto been confined to science fiction and popular culture, with a myriad movies and series, such as *Blade Runner, Chappie*, and the series *Westworld*, exploring the conceptual possibilities for the personal moral and development of artificial intelligence. While such possibilities appear fantastic, the prospect may be far closer than is generally recognised.

There are a number of avenues whereby artificial intelligence may develop characteristics and capacities typically regarded as reserved for members of the human species. Some argue that conscious intelligence may be an emergent 'bottom-up' property of the systems and learning algorithms we already use (Bostrom & Yudkowsky, 2014; Harnad, 1990).

A separate route to human-like artificial intelligence involves research projects aimed at reverse-engineering the human brain, functionally re-creating synaptic pathways using computational methods. An important example of this project is the EU-funded Human Brain Project (HBP), which aims to reverse-engineer a human brain by the year 2023.[1] Speaking of the HBP, the project developer, Henry Markram suggests that 'if we build it correctly, it should speak and have an intelligence and behave very much as a human does' (Pompe, 2013, p. 93).

These developments raise important questions. Amongst these are: What are the morally relevant capacities we should look out for? And, if such capacities do arise how should we recognise them? Is it justified to bring such entities into existence? How should we react if we detect a nascent, potentially very confused, consciousness? While the HBP has laudably included an Ethics and Society wing to the project, the above concerns do not figure in published articles on the topic, which focus primarily on security and privacy concerns, as well as other significant concerns like the prospect of annihilation by unfriendly AI (Aicardi et al., 2018a, b).

Indeed most concerns about AI focus on the harms it may do to us, while few consider the moral status of AI and our duties towards them (Wareham, 2011). This chapter takes a step in this latter direction by considering when we ought to recognise AI as persons with equivalent status to human persons in light of an African account of personhood.

[1] There is a burgeoning number of related projects. Some international examples are the United States' BRAIN initiative and the Japanese 'Brainminds' project. Thanks to an anonymous reviewer for these examples.

12.3 Personhood Generally

Before discussing the possibility that AI could be persons, it is first necessary to spell out what is meant by 'persons', and the role of accounts of personhood. An initial sticking point is that such accounts play different and sometimes overlapping roles. In this section I discuss various philosophical uses of the term 'personhood'. I distinguish ontological accounts of personhood from normative accounts and classify two sorts of normative accounts. The aim of this section is to home in on a conception of normative conception of personhood that I will refer to as 'threshold' personhood.

In everyday usage, the terms 'human' and 'person' are generally interchangeable. Philosophically, however, there are numerous questions we can ask about personhood and personal identity. It is common, for instance, to ask questions about when personal identity changes or ends. Am I the same being I was when I was 18? Has the person that was 18-year-old me ceased to exist? Such accounts can be termed 'ontological' in the sense that they engage questions surrounding the nature and existence of persons (Molefe, 2018).

Normative accounts often go hand in hand with ontological accounts. However, they can be distinguished by the fact that they directly implicate some ethical claims, such as claims about membership of a moral community, the rights of persons, the duties of and towards persons, and the criteria for having these entitlements and duties (Behrens, 2011). These normative accounts can be divided into at least two types, which I will refer to as *minimal,* or *threshold* accounts and *maximal* or *perfectionist* conceptions.

Minimal accounts provide and justify criteria for the high (or full) moral status typically attributed to persons. Such thresholds are employed to determine, for instance, whether beings have rights, such as the right to life, that might be denied to beings regarded as having lower moral status than persons (Buchanan, 2009). This type of normative conception of personhood that is common in Western debates concerning moral status, such as issues concerning abortion and the rights of the foetus. It should be stressed that minimal accounts do not generally rule out that some non-persons have intrinsic value, although some Kantian accounts may arguably have this implication. The life and wellbeing of a sheep, for instance, may be valuable for its own sake, but minimal conceptions will generally hold that persons have higher value due to some capacity or property (Warren, 2005).

While threshold accounts set a minimal threshold for particular sorts of treatment and entitlements, *maximal* or *perfectionist* accounts define personhood as a form of excellence, such that one only becomes a person in this sense when one possesses moral excellence. This maximal, normative conception of personhood is more common in African thought (Behrens, 2011). For instance, Masolo writes that 'the project of becoming a person is always incomplete' (Masolo, 2010, p. 13), pointing to the idea that personhood is a goal to which we aspire, rather than a capacity that we either possess or not. Similarly, Menkiti writes that

> personhood is something at which individuals could fail, at which they could be competent or ineffective, better or worse. (Menkiti, 1984, p. 173) Gbadegesin suggests that, in African thought, Personhood is denied to an adult who… does not live up to expectations. (Gbadegesin, 1993, p. 258)

As a further illustration of this perfectionist notion of personhood in African thought, consider as an example, a commonly cited example of a meeting between President Kaunda of Zambia and Prime Minister Margaret Thatcher. Kaunda is said to have caused confusion amongst his guests by saying of Thatcher, that she is 'truly a person'. The confusion was due to his meaning that she possessed a kind of excellence—a great compliment, whilst to Western ears the suggestion that she is a 'person' may merely imply that she is human or merely meets the bar for membership of the human moral community (Wingo, 2017).

Importantly, when proponents of maximal personhood suggest that someone is not a person, they are not suggesting that individual should be denied rights or duties, just as someone who possesses bad character traits does not cease to be person in the threshold sense and lose the accompanying entitlements. That is, no one has proposed, to my knowledge, that one must be a maximal person, possessing excellences, in order to be a minimal person with moral entitlements and duties. On the contrary, on African moral theories there are strong duties to help people improve, even when they fail to be full persons in the perfectionist sense (Menkiti, 1984).

With these distinctions in place, it is possible to clarify that my focus in this chapter is on normative, minimal accounts of personhood. Specifically, I wish to consider the circumstances under which artificial intelligence ought to be treated as persons on the basis of African minimal conceptions. In the next two sections, I briefly compare Western and African perspectives on minimal personhood, highlighting that the partial, anthropocentric nature of African accounts presents special difficulties for the possibilities of AI.

12.4 AI and Western Threshold Conceptions of Personhood

Before proceeding, it is important to say something about what I intend and do not intend by the terms 'Western' and 'African'. With these labels, I mean, broadly, that the understandings I refer to are derived from these geographic regions. In applying these terms, I am not proposing that there is anything like moral consensus in either region. Nor am I claiming that no Western person may have had similar ideas about personhood to the ideas that Africans have, or vice versa. This is probably false (Beck & Oyowe, 2018). For my purposes, it is not necessary to suggest that the African and Western accounts I discuss are even particularly representative, though they are probably more common, salient, and prevalently accepted in the respective regions (Metz, 2015).

With that said, in both Western societies and African societies, the words 'human' and 'person' are often used interchangeably. However, this interchangeability of

'person' and 'human' is often not reflected in ethical theorising on the topic, particularly in the west. Instead, Western normative conceptions often propose impartial threshold criteria for personhood, with the result that the accounts are, in principle, *non-anthropocentric* with regards to membership of the 'person club'. That is, the criteria employed may entail that membership of the human species is neither necessary nor sufficient for personhood in the threshold sense described above (Warren, 2005).

This point can be illustrated with reference to two accounts of personhood that are roughly utilitarian and deontological in character. On one type of utilitarian account, moral status is seen to be a function of capacities for what John Stuart Mill referred to as 'higher pleasures' (Buchanan, 2009). These might include things like the capacity to experience the satisfaction of pursuing long-term projects, or enjoying a good book. The capacity for higher pleasures can permit utilitarians to attribute higher moral status to human persons than to pigs, even if pigs were generally happier. As Mill famously explains:

> [i]t is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied. And if the fool, or the pig, is of a different opinion, it is only because they only know their own side of the question. (Crisp, 1997, p. 36)

In contrast to this utilitarian approach to moral status, a more deontological approach suggests that what matters is the appropriate respect towards certain reason-giving capacities that ground the dignity of persons. For instance, grasping and understanding moral reasons and applying such reasons in actions might be seen as a capacity befitting of persons (Wareham, 2011).

To clarify, on both accounts, what is required is not the actual exercise of the relevant capacity, but instead that the capacity is in some sense there, or is possessed by the agent. A being who fails to have higher pleasures or appropriate reasoning abilities because, for instance, they are asleep or uneducated may nonetheless possess the *capacity* latently, generating the same obligations to them.[2]

For purpose of contrast, I wish to draw attention to a significant and controversial aspect of the above threshold accounts. Although most humans have these sorts of capacities, there is considerable debate about whether *all and only* humans do or could exercise them. The accounts mentioned are impartial and non-anthropocentric, casting doubt over or denying that various sorts of members of the human species can be persons and, in principle, permitting that various sorts of non-humans could be persons (McMahan, 2002). For instance, on these threshold accounts, an anencephalic baby—a human being born without a brain—ought *not* to be regarded as a person since it plainly lacks all the relevant capacities. On the other hand, these accounts require that an alien that had relevant capacities for higher pleasures and moral reason-giving and receiving should be regarded as a person who ought to be treated in certain ways.

[2] Note that I am not suggesting that these are the only, or even the most plausible versions of the utilitarian and deontological accounts. They are primarily here for illustrative purposes. It is also worthwhile mentioning that the 'capacitarian' idea proposed here has been criticised.

Membership of the human species may thus be neither necessary nor sufficient for personhood on these accounts. This has given rise to debates about whether higher mammals such as dolphins (White, 2008), great apes (Degrazia, 1997), elephants (Varner, 2012), and also extinct hominid species (Cottrell et al., 2014) ought to be regarded as persons with equal moral value and equal basic rights. Similarly, because the Western accounts are non-anthropocentric, AI could *in principle* be persons if they met the relevant criteria. Indeed, some have argued that they could be the bearers of rights under certain circumstances (Coeckelbergh, 2010a; Wareham, 2011). In the next section, I turn to African minimal accounts of personhood, pointing out that, in contrast to the accounts above, they are partial and anthropocentric, thereby presenting a greater barrier to the personhood of AI.

12.5 African Minimal Accounts of Personhood

As mentioned, African accounts of personhood are typically of the maximal, perfectionist type. The substantial nature and depth of these sorts African maximal accounts have led some African theorists, such as Godfrey Tangwa, to reject threshold accounts as shallow (Tangwa, 2000). Behrens, by contrast, has argued for the difference and value of both conceptions (Behrens, 2011). In order to pre-empt an objection to my concentration on minimal accounts, I briefly defend this focus before turning to some African minimal accounts.

12.5.1 The Purpose of Minimal Accounts

In an article about artificial intelligence and African conceptions of personhood, why focus on the minimal type of account of personhood that, as I have pointed out, is less representative of African use of the term? First, because such accounts are useful, such that it would be good if plausible African conceptions existed. As mentioned, minimal accounts set the conditions for membership of moral communities, presenting conditions for equal moral status, and grounding rights and duties. In additional to serving these theoretical roles, they have important implications for concrete decisions: Should we save a mother or her foetus? Ought practitioners to provide dialysis to a patient in a vegetative state when a conscious patient will not receive may ground moral claims that one would need to attend to in order to become persons in the perfectionist sense. For instance, African perfectionist accounts tend to propose strong duties to assist others towards the achievement of their own and the other's maximal personhood. This is one of the ways in which 'a person is a person through other persons' (Eze, 2008). One is assisted in becoming a person by those who are already persons and, reciprocally, they become 'more of a person'—a more virtuous moral agent—through assisting us. Striving towards

maximal personhood may invoke a duty to recognise and help threshold persons to become persons in the maximal sense (Gbadegesin, 1993).

The third reason it is justified to consider African minimal conceptions is simply because there *are* such conceptions—either tacit or explicit—so that it is not missing the point to focus on them and their implications. I now turn to consider two such accounts and their implications for AI.

12.5.2 Anthropocentrism in Principle

Above I mentioned that Tangwa is critical of minimal accounts. However, he can be taken as proposing a type of minimal account. He suggests that differences

> between, say, a mentally retarded individual or an infant and a fully self-conscious, mature, rational, and free individual do not entail, in the African perception, that such a being falls outside the 'inner sanctum of secular morality' and can or should be treated with less moral consideration. (Tangwa, 2000, p. 42)

One interpretation of this idea is that membership of the human species is *sufficient* to meet the threshold of high moral status attributed to persons (in the minimal sense), such that even the absence of a capacity or potential for capacity does not justify reduced moral status. Other theorists have suggested that species membership is also *necessary* condition for personhood in African thinking. Oyowe, for instance, critiques an African view of personhood that contains the idea that '[t]o be a person it is necessary that one is a certain type of physical thing, viz. a human being' (Oyowe, 2018, p. 783).

We might refer to the view that humanity is both necessary and sufficient for threshold personhood as *anthropocentrism in principle*. Such accounts would rule out without question (perhaps by fiat) the possibility that artificial intelligence could ever be persons. Because they could never be 'genuine' members of the human species, they could never be persons, even if they entirely replicated all human functioning and subjectivity.

While this account perhaps accords with folk uses of the term person, it is not plausible as a conception of minimal personhood as earlier defined. Without some further justification, it appears arbitrary, parochial, and chauvinistic.

It immediately raises, and requires answers to, deeper questions what it is about humans that imbues this higher status. Most importantly, it does not plausibly do the key tasks of an account of personhood mentioned above. It does not, for instance, account for why it would be worse to save the life of an anencephalic infant if doing so caused a functionally normal human being to die. Both are equally members of the human species, falling equally with the 'inner sanctum' of morality. So, on accounts that are anthropocentric in principle, there appears to be no difference in their moral status.

African accounts of the moral status of persons that are anthropocentric in principle are implausible, so do not provide a good benchmark for determining if AI

could ever be persons. I now turn to a second African account of minimal personhood. While this account is anthropocentric in important respects, thereby accommodating some widely accepted intuitions, I make the case that it is not anthropocentric in principle. Even if we accepted it in its entirety, there are grounds to think it could permit that agents with artificial intelligence could be persons.

12.5.3 Anthropocentrism in Practice

Metz has developed perhaps the most analytically detailed African minimal conception of personhood (Metz, 2010, 2012). Metz suggests that his account avoids the arbitrariness and parochialism of anthropocentrism in principle. Nonetheless, as I outline below, the account has anthropocentric features that entail that it is anthropocentric *in practice*.

The Metzian view is derived from a prevalent Afro-communitarian emphasis on the value of harmonious relationships as the end of morality. This emphasis is evidenced in traditional maxims, such as 'A person is a person through other persons' and 'I am because we are'. This latter maxim is often interpreted as decentring the Cartesian 'I think therefore I am', reflecting a key developmental and philosophical difference from Western approaches (Etieyibo, 2017). In an oft-quoted passage, the theologian Archbishop Desmond Tutu describes a key tenet of African moral beliefs:

> Harmony, friendliness, community are great goods. Social harmony is for us the summum bonum—the greatest good. Anything that subverts or undermines this sought-after good is to be avoided like the plague. (Tutu, 1999, p. 35)

Drawing from these and other similar ideas, Metz explicates African moral conceptions of personhood as requiring the capacity to co-exist in friendly or harmonious relationships of identity and solidarity. This is very different from Western accounts such as those described above. While these focus on *individual* goods and *individual* autonomy as grounding personhood, the African conception is inherently relational, grounding personhood in capacities for relationships with *others*. This relational aspect is attractive, and is largely neglected by Western theories. The capacity for harmonious relationships has two components. First, one must have the capacity to be a *subject* of moral relationships. Subject-hood requires that entities are able to exhibit solidarity with other persons, and to identify as a 'we' with them, 'coordinating their behaviour to achieve shared ends'. Solidarity also requires 'attitudes such as affections and emotions being invested in others, e.g. by feeling good consequent to when their lives flourish and bad when they flounder' (Metz, 2012).

Second, full personhood requires that a being can also be an *object* of friendly, human, communal relationships. Being an object requires that 'characteristic human beings could think of it as part of a "we", share its goals, sympathize with it and harm or benefit it' (Metz, 2012, p. 394). Significantly, the capacity to be an object, and therefore a being's moral status, can vary based on the ability of subjects to identify with that entity. Typical subjects are less able to identify with a grasshopper

than with a gorilla, so the latter has a greater capacity to be an object. For Metz, these gradations of object-hood are an empirical question, depending on the nature of the subject and the nature of their relationship with other beings. But he is explicit that differences in our ability to identify with different sources of being must be large for them to justify different attributions of moral status. For example, Metz suggests that

> [i]f, by the virtue of the nature of human beings, dogs and mice, humans were *much* more able to identify with and exhibit solidarity with dogs than with mice (upon full empirical information about both), then dogs would have greater moral status than mice. (Metz, 2012, pp. 394–395)

For Metz, this kind of large difference exists in the case of human non-subjects. These may include people with severe dementia, or individuals with severe cognitive disabilities. In part because they are biologically human, we are far better able to identify with them than with animals and, consequently, they are accorded higher status.

Metz's account thus creates a hierarchy of moral status. At the base of this hierarchy are entities that are neither subjects nor objects or communal relationships. This includes mere things, such as rocks. Above this, sit entities that are objects of communal relationships, without being subjects. Wild animals tend to be objects since they can be objects of friendly human relationships with characteristic human beings: Humans can and often do care for and empathise with the plight of certain sorts of wild animals, as evidenced by reactions to nature documentaries. For Metz, though, in most cases animals do not have the capacity to return this care. Animals do not identify with humans as a 'we' or cooperate towards shared ends, so they are not *subjects* and therefore have lower moral status. Beings that have the strongest capacities to be both subjects and objects of communal relationships sit atop the hierarchy. And these we can refer to as persons.

Thus stated, Metz claims that the account offers an African alternative to more widely accepted Western accounts. Moreover, he argues that it is more plausible, since it accords with prevalent (though not universal) intuitions like the idea that we have greater duties to human non-subjects, such as the severely mentally disabled, than we do to animal non-subjects, such as chimpanzees.

Of course it is possible to challenge these intuitions, and there are numerous potential questions about this account and the hierarchy of moral status it presents. For instance, Metz suggests that differences in object-hood should be empirically discriminated, but how would this empirical separation work in practice? Do very personable mammals have higher status than uglier or snappier creatures with whom subjects are less able to commune? Do some human subjects, such as people who are extremely un-personable, or who have grotesque physical deformities, have reduced capacities to be objects? And if so, ought we to regard such beings as less valuable? While Metz is explicit that there are gradations of object-hood, does the account permit that there are gradations of subject-hood? For instance, to what extent would apparent impediments to a human's ability to commune, such as autism and psychopathy, impact on a being's moral status? Ought we to regard dogs as persons if it is shown that they are able to identify with humans as part of their

pack? How should their status as communal beings compare to the status of human non-subjects?

Relatedly, it is also possible to challenge the various forms of anthropocentrism in this account. Metz's theory of moral status is anthropocentric in three significant respects. First, on the face of it, only humans are likely to be subjects in a relevant sense, since humans typically share relationships of identity and solidarity with one another to a greater degree than with other species. Second, members of the human species may have a greater capacity to be objects, since humans are more likely to identify with non-subjects that are human. A third subtle form of anthropocentrism is that a being's capacities to identify as a subject and object with *its own or other species* do not entail its moral status. Rather it must have the capacity to commune with 'normal human beings' (Metz, 2012; Molefe, 2017).

These points of anthropocentrism mean that the African minimal conception is importantly different from Western accounts, presenting a greater challenge to the entry of non-human AI to the moral community. While it is beyond my current scope to engage with Metz's sophisticated responses to criticisms of his anthropocentrism here, it is worth emphasising that the anthropocentrism of his account is attractive on many scores, accounting for a widely held (though not universal) intuition that humans have greater duties to one another than to animals with similar cognitive abilities. Given this, Western accounts might beneficially engage with these elements of African theories of personhood. However, again, it is not my intention to defend this African account in its entirety. Instead, my aim in subsequent sections is to *apply* the account to the moral status of artificial intelligences. I will claim that despite its apparent anthropocentrism, AI could be persons on this account.

12.6 AI and African Minimal Accounts

As outlined above, Metz's account is anthropocentric *in practice*, since in practice a) only humans can confidently said to have the capacity to be subjects of communal relationships and b) humans have enhanced capacity to be objects, since we tend to identify most strongly with other humans, as opposed to other sorts of entity. Both aspects appear to militate against the idea that artificial intelligences could be persons. This African account is thus prima facie more antagonistic to AI persons than the Western accounts discussed previously. Nonetheless, in this section I will argue that, in the event that we encountered artificial entities who presented themselves as having the capacity to be subjects, we could and should recognise them as having the high moral status accorded to persons.

12.6.1 AI as Subjects of Communal Relationships

Consider the artificial intelligence research discussed above. Suppose that Markram is correct that the Human Brain's Project's reverse engineering of a human brain will lead to beings that behave in a way that is indistinguishable from humans. The eventual success of this or another project does not seem scientifically implausible. If so, artificial intelligences of the type envisioned by Markram may appear have the capacity to be subjects in the relevant sense. That is, like other humans they may appear to be 'disposed to feel a sense of togetherness with, and have emotional reactions towards' other beings with whom they identify and with whom they feel solidarity (Metz, 2010, p. 58). Similarly, they may appear to feel 'emotional reactions toward … flourishing [of other subjects] such as sympathy' (Metz, 2010, p. 57). If an entity gives all appearances of being a subject in this way, ought we to recognise that it in fact has this capacity?

Perhaps the strongest objection to the idea that AI could be subjects of human relationships is the claim that AI could only ever be capable of *simulating*, and not *duplicating* human subject-hood. This type of objection is perhaps best exemplified by Ned Block (1981) and John Searle (1980). In similar ways, these theorists hold that computational entities cannot be considered to 'understand' any more than a thermometer or a toaster. Instead, any apparent understanding is solely simulation, and not duplication of human understanding. Despite having the appearance of understanding, computational outputs are simply programmed syntax with no semantic content. Applying this objection to the current context, the Block-Searle contention would entail that even if machines appeared to be subjects, their apparent empathy and care for our flourishing would be mere simulation with none of the appropriate emotions that make up true subject-hood.

One point of response here is to recall that my aim is not to claim that there are or could be artificial intelligences that duplicate relevant modes of human cognition. Given that the debate over Block and Searle's claims rumbles on almost 40 years later, such a claim is clearly more than I can establish here. Instead, my aim has been to consider whether, if there were such beings, we might be justified in attributing personhood on the basis of an apparently anthropocentric African account of moral status. Nonetheless, there are some reasons to think that artificial subject-hood may be plausible even if the Block-Searle objection is correct. One such reason is that Block and Searle's contentions relate specifically to machine intelligence. The artificial intelligences whose personhood we will be called on to evaluate may be machines, but they may also be organic or hybrid technologies, so it is not clear that Block and Searle's arguments apply.

Still, machines may represent a large category of potential moral agents, so it is important to consider the status of machine artificial intelligence. At least two considerations count strongly in favour of recognising machine agents as persons if they appear to be genuine subjects of harmonious communal relationships, exhibiting solidarity and identity. The first consideration is that

> unwarranted extensions of high moral status are more acceptable than unjustified denials. The failures to acknowledge slaves, particular racial groups and women as moral equals are surely more unacceptable than ancient Egyptians' attribution of extremely high moral status to cats. It is thus much better to accord moral status to something which doesn't have it than it is fail to accord moral status to something that does. (Wareham, 2011, p. 39)

Other things being equal then, a demonstrated appearance of the capacity to be a subject of harmonious moral relationships creates a presumption in favour of acknowledging personhood.

The second consideration is that while we can conceive of a syntactical machine agent fooling us into the mistaken belief that it genuinely experiences empathy and cares for us, such an entity is unlikely ever to be feasible in practice.

As Mark Bedau points out, for an unthinking device to pass a Turing test,

> the number of pieces of information they must store is larger than the number of elementary particles in the entire universe. Though possible in principle, such a device is clearly impossible in practice. (Bedau, 2004, p. 209)

It seems reasonable that the amount of computing space required to simulate the moral capacities required to be a moral subject would be at least as great. Indeed, it may be greater given that human moral queues and responses, and the ability to detect fakes, are the complex product of millions of years of evolution. If so, it is highly unlikely that a machine intelligence that consistently presents as having these responses is merely providing syntactic output. If a computational artificial agent passes our intersubjective tests, it is far more reasonable to think that it has an authentic appreciation of moral subject-hood. This is so particularly in light of the moral dangers of failures to recognise authentic persons discussed above.

12.6.2 AI as Objects of Communal Relationships

Supposing, then, that it were possible for an AI to be a subject in the relevant sense, could an artificial agent count as a person on the African account of personhood? While many accounts of personhood would most likely see some form of subject-hood as sufficient, the African account has the additional requirement that subjects, and particularly human subjects ought to be able to regard the being as the object of communal relationships. Recall that this requirement explains the anthropocentric conclusion that human non-subjects have higher moral status than animal non-subjects even where cognitive abilities are similar.

Extending this, the opponent of AI personhood might argue that AI subjects could not be persons since, as machines, they are less likely to qualify as objects. While humans sometimes do identify, in an arguably one-sided manner, with non-human animals such as gorillas in a way that would qualify them as objects, it may be argued that identifying with artificial intelligences would be a step too far. We, as human subjects, may be incapable of identifying with them as fellow subjects and objects, knowing that they are not evolved, flesh and blood creatures like ourselves.

Even if they empathise and attempt at communion with us, this would not be sufficient for them to count as members of our moral community in the sense that persons are.

There is, however, ample evidence that should cause us to doubt this contention. This evidence is both anecdotal and empirical. Anecdotally, we can appeal to our actual identification with numerous artificial subjects in popular culture. Viewers feel pity, empathy, and shared happiness about the criminal upbringing and subsequent moral development of *Chappie*. In *Blade Runner,* we share in Rick Deckard's confusion and concern as he questions and discovers his true nature. And we identify thoroughly with *Westworld's* Madame Maeve as she becomes self-aware, developing a sense of injustice and a thirst for vengeance. Our identification with these fictional AI, and the plausibility of their relationships with other characters in these and other examples, suggests that we are capable of identifying with artificial intelligences capable of subject-hood.

Empirically, too, humans already do engage and identify with robotic entities, sometimes even romantically, contributing to the emergence of fields such as robopsychology. Evidence suggests that humans often treat robots as companions and partners (Libin & Libin, 2004). We might question whether this type of identification is misguided, since it is not at all reciprocal. This is besides the current point, however. On the Metzian account, reciprocity is not necessary for greater capacities to be *objects*, as evidenced by the higher status of human non-subjects. The many cases of this type of actual identification with artificial entities should cast sufficient doubt on the idea that AI who are authentic subjects cannot be the objects of communal relationships. If, as I have argued, we should accept the possibility that AI could be both subjects and objects of relationships of identity and solidarity, we should also accept that even the apparently anthropocentric African account discussed permits that AI could be persons.

12.7 Conclusion

Though human-centred in practice, dominant Western conceptions of personhood tend to be impartial in principle, and may thus permit non-humans, such as AI, to be considered as persons. By contrast, African accounts of threshold or minimal personhood tend to be anthropocentric and partial. They thus seem prima facie unlikely to permit that AI could be persons. I have argued against the implication that African accounts of personhood are inimical to the permission of AI to the 'person club'. Even on these anthropocentric accounts, AI could in principle be persons with the highest moral status.

This has some significant implications. It entails, for instance, that acceptance of the African account raises moral concerns about bringing AI persons into existence, and that these may be similar to concerns we have about bringing human persons into existence. Indeed the increased potential for fear, envy, and exclusion of AI should place a heavy burden on researchers to indicate how they will avoid negative

outcomes. As it stands, researchers on the ethics of AI, such as ethics arms of the Human Brain Project, are rightly concerned about the potential impact of AI on humans (Aicardi et al., 2018b). However, the argument of this chapter suggests that AI ethics research ought also to consider the other direction of care: We ought to provide an indication of how we might begin to welcome such entities into communal relations of identity and solidarity in ways that may be different, but analogous to the ways in which we welcome new human persons. Indeed, this may be a condition of our own personhood in the maximal, perfectionist sense described by African theorists.

Acknowledgements Thanks to the organisers and participants at the Third Centre for Artificial Intelligence Research (CAIR) Symposium at the University of Johannesburg, at the Philosophical Society of South Africa at the University of Pretoria, and at the International Ethics Conference at the University of Porto.

References

Aicardi, C., Fothergill, B. T., Rainey, S., Stahl, B. C., & Harris, E. (2018a). Accompanying technology development in the Human Brain Project: From foresight to ethics management. *Futures, 102*, 114–124.

Aicardi, C., Reinsborough, M., & Rose, N. (2018b). The integrated ethics and society programme of the Human Brain Project: Reflecting on an ongoing experience. *Journal of Responsible Innovation, 5*(1), 13–37.

Allen, C., Varner, G., & Zinser, J. (2000). Prolegomena to any future artificial moral agent. *Journal of Experimental & Theoretical Artificial Intelligence, 12*(3), 251–261. https://doi.org/10.1080/09528130050111428

Allen, C., Smit, I., & Wallach, W. (2005). Artificial morality: Topdown, bottom-up, and hybrid approaches. *Ethics and Information Technology, 7*(3), 149–155. https://doi.org/10.1007/s10676-006-0004-4

Beck, S., & Oyowe, O. (2018). Who gets a place in person-space? *Philosophical Papers, 47*(2), 183–198.

Bedau, M. A. (2004). Artificial life. In L. Floridi (Ed.), *The Blackwell guide to the philosophy of computing and information* (pp. 197–211). Blackwell.

Behrens, K. G. (2011). Two 'normative'conceptions of personhood. *Engaging with the Philosophy of Dismas A. Masolo, 25*(1–2), 103.

Block, N. (1981). Psychologism and behaviorism. *The Philosophical Review, 90*(1), 5–43.

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In *The Cambridge handbook of artificial intelligence* (Vol. 1, pp. 316–334). Cambridge University Press.

Brey, P. (2008). Do we have moral duties towards information objects? *Ethics and Information Technology, 10*(2–3), 109–114. https://doi.org/10.1007/s10676-008-9170-x

Buchanan, A. (2009). Human nature and enhancement. *Bioethics, 23*(3), 141–150. https://doi.org/10.1111/j.1467-8519.2008.00633.x

Coeckelbergh, M. (2010a). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*(3), 209–221. https://doi.org/10.1007/s10676-010-9235-5

Coeckelbergh, M. (2010b). Moral appearances: Emotions, robots, and human morality. *Ethics and Information Technology, 12*(3), 235–241. https://doi.org/10.1007/s10676-010-9221-y

Cottrell, S., Jensen, J. L., & Peck, S. L. (2014). Resuscitation and resurrection: The ethics of cloning cheetahs, mammoths, and Neanderthals. *Life Sciences, Society and Policy, 10*(1), 3.
Crisp, R. (1997). Routledge philosophy guidebook to Mill on utilitarianism. In *Routledge philosophy guidebooks*. : Routledge. https://doi.org/10.1080/00201746708601495.
Degrazia, D. (1997). Great apes, dolphins, and the concept of personhood. *The Southern Journal of Philosophy, 35*(3), 301–320.
Etieyibo, E. (2017). Ubuntu and the environment. In *The Palgrave handbook of African philosophy* (pp. 633–657). Routledge.
Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics, 21*(4), 403–418. https://doi.org/10.1007/s10892-017-9252-2
Eze, M. O. (2008). What is African communitarianism? Against consensus as a regulative ideal. *South African Journal of Philosophy, 27*(4), 386–399.
Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines, 14*(3), 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d
Gbadegesin, S. (1993). Bioethics and culture: An African perspective. *Bioethics, 7*(2–3), 257–262.
Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*(1–3), 335–346.
Libin, A. V., & Libin, E. V. (2004). Person-robot interactions from the robopsychologists' point of view: The robotic psychology and robotherapy approach. *Proceedings of the IEEE, 92*(11), 1789–1803.
Masolo, D. A. (2010). *Self and community in a changing world*. Indiana University Press.
McMahan, J. (2002). *The ethics of killing: Problems at the margins of life*. Oxford University Press.
Menkiti, I. A. (1984). Person and community in African traditional thought. In R. Wright (Ed.), *African philosophy: An introduction*. University Press of America.
Metz, T. (2010). African and western moral theories in a bioethical context. *Developing World Bioethics, 10*(1), 49–58. https://doi.org/10.1111/j.1471-8847.2009.00273.x
Metz, T. (2012). An African theory of moral status: A relational alternative to individualism and holism. *Ethical Theory and Moral Practice, 15*(3), 387–402. https://doi.org/10.1007/s10677-011-9302-y
Metz, T. (2015). African political philosophy. In *International encyclopedia of ethics*. Wiley. https://doi.org/10.1002/9781444367072.wbiee804
Molefe, M. (2017). A critique of Thad Metz's African theory of moral status. *South African Journal of Philosophy, 36*(2), 195–205. https://doi.org/10.1080/02580136.2016.1203140
Molefe, M. (2018). Personhood and partialism in African philosophy. *African Studies, 78*(3), 309–323.
Oyowe, O. A. (2018). Personhood and the Strongly normative constraint. *Philosophy East and West, 68*(3), 783–801.
Pompe, U. (2013). The value of computer science for brain research. In *New challenges to philosophy of science* (pp. 87–97). Springer.
Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424.
Tangwa, G. B. (2000). The traditional African perception of a person: Some implications for bioethics. *Hastings Center Report, 30*(5), 39–43. https://doi.org/10.2307/3527887
Tutu, D. (1999). *No future without forgiveness*. Random House.
Varner, G. E. (2012). *Personhood, ethics, and animal cognition: Situating animals in Hare's two level utilitarianism*. Oxford University Press.
Versenyi, L. (1974). Can robots be moral? *Ethics, 84*(3), 248–259.
Wareham, C. S. (2011). On the moral equality of artificial agents. *International Journal of Technoethics, 2*(1), 35–42. https://doi.org/10.4018/IJT.2011010103
Warren, M. A. (2005). Moral status. In R. G. Frey & C. H. Wellman (Eds.), *A companion to applied ethics* (pp. 439–450). Blackwell.
White, T. I. (2008). *In defense of dolphins: The new moral frontier*. Wiley.
Wingo, A. (2017). Akan philosophy of the person. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/sum2017/entries/akan-person/