



Artificial General Intelligence and Panentheism

Oliver Li

To cite this article: Oliver Li (2023) Artificial General Intelligence and Panentheism, Theology and Science, 21:2, 273-287, DOI: [10.1080/14746700.2023.2188373](https://doi.org/10.1080/14746700.2023.2188373)

To link to this article: <https://doi.org/10.1080/14746700.2023.2188373>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 20 Mar 2023.



Submit your article to this journal [↗](#)



Article views: 1146



View related articles [↗](#)



View Crossmark data [↗](#)



Artificial General Intelligence and Panentheism

Oliver Li

ABSTRACT

In this article, I argue that given the possibility and prospect of Artificial General Intelligence (AGI), panentheism, as a form of theism with a stronger emphasis on the immanence of God, parallels the anti-anthropocentrism implied by AGI. I discuss some general issues related to the categorization of Artificial Intelligence (AI). Next, both anthropomorphism and anthropocentrism will be discussed as concepts for how humans may relate to AI. Subsequently, I argue and conclude that there is an analogy between the anti-anthropocentric implications of AGI and the anti-anthropocentric element of panentheism, but that panentheism points to a stronger form of anti-anthropocentrism.

KEYWORDS

Artificial General Intelligence; panentheism; anthropocentrism; anthropomorphism

Introduction

Since the development of computers, questions concerning AI (artificial intelligence) have been thematized in popular culture (for example, in science-fiction literature and screenplays), and discussed in various academic disciplines such as computer science and philosophy of mind. The breakthrough of computer technology into society in the '80s, together with the accelerating development, which followed Moore's law and led to more advanced and powerful applications, has made questions concerning AI ever more pressing. It may be easy to imagine that the abilities of computer systems will soon surpass human abilities, not only in specific areas such as playing Go or handling Big Data, but also more generally. This raises the following question, which has both a metaphysical and practical character: Is it possible for an advanced computer system to have a mind, consciousness, a conscious mind, or even feelings? Computer scientists and software developers usually work with the practical side of this question, while philosophers are interested in the metaphysical aspects of this question. Irrespective of whether this question is answered in the affirmative in practice or as a metaphysical question, one may reflect on the consequences that the creation of Artificial General Intelligence (AGI) would bring for the relation of a human, and a religious human in particular, to the world in general? Focusing on a more specific example of theism, namely panentheism, as an example of a religious worldview, I will argue and conclude that given the possibility and prospect of AGI, panentheism, with its stronger emphasis on the immanence of God, leads to similar consequences as the possible development of AGI, yet takes them one step further.

In a first step, I discuss some general issues related to the categorization of AI and the possibility of AGI. Further, both anthropomorphism and anthropocentrism will be identified as important concepts for understanding how humans may perceive AI and relate to AI in general and AGI in particular. Subsequently, I argue and conclude that given the implications of AGI for anthropocentrism and anthropomorphism, there is an analogy between the consequences for human relations of developing AGI concerning anthropomorphism and anthropocentrism, and a panentheistic worldview. Furthermore, I emphasize that the non-anthropocentric consequences of developing or acknowledging the possibility of AGI and the implications of panentheism need not be interpreted as dehumanizing and that panentheism leads to a stronger form of non-anthropocentrism.

Preliminaries on the Categorization of AI

Initially, one may reflect on the question of what is meant by the term “Artificial Intelligence.” In the following, I briefly present some examples of how scholars have defined or described AI. John R. Searle distinguishes between strong-AI and weak-AI. Regarding the former, he writes: “On the Strong AI view, the appropriately programmed digital computer does not just simulate having a mind, it literally has a mind.”¹ Weak-AI would then entail simulations of the mind or parts of the mind. According to David J. Chalmers, “the field of *artificial intelligence* (or AI) is devoted in large part to the goal of reproducing mentality in computational machines.”² Manuela Lenzen distinguishes three approaches in the field of AI: (a) the development of systems which perform specific tasks, (b) to understand human cognition by understanding artificial systems, and (c) to develop AGI, understood as intelligence which has the full flexibility and universality of the mind.³ The latter approach is also what Ted Peters describes as the goal of the strong AI movement.⁴ With a stronger focus on religious aspects, Noreen Herzfeld introduces the term *imago hominis*: “Thus, AI can be viewed as the attempt to create an *imago hominis*, a machine that is in some way created in the image of the human person [...]”⁵ John McCarthy simply talks about human-level AI, while Andreas Kaplan and Michael Haenlein distinguish between (a) artificial narrow intelligence (ANI) which applies AI to specific areas, (b) AGI which applies AI to several areas, is autonomous, and equals humans in several areas, and (c) artificial superintelligence (ASI) which would apply to and outperform humans in any area.⁶ By distinguishing ANI, AGI, and ASI, Kaplan and Haenlein wish to capture the evolutionary aspect of moving from ever more advanced ANI systems to AGI and eventually to ASI. In this article, I will follow the final distinction by Kaplan and Haenlein, which firstly captures the common distinction between weak-AI and strong-AI, but which additionally allows for the possibility that an AI in the future will surpass the abilities of humans in any field. Interestingly, Kaplan and Haenlein nevertheless believe without further argument or motivation that human abilities such as creativity and emotions most probably cannot be realized in an AI-system.⁷

One crucial observation is that all of the above descriptions or definitions in some way or another are based on *human* intelligence. Although this observation may seem trivial – after all, humans tend to compare other beings with themselves – I wish to emphasize this observation since in a sense it places humans at the center of the description or definitions. In fact, in a review of tests for machine consciousness, Aida Elamrani and

Roman V. Yamploskiy find it “striking” that “*human* evaluation is *always* required at some point” (my emphasis) in tests for consciousness.⁸

Furthermore, all of the above descriptions or definitions call for methods or tests for how to decide whether a machine should be regarded as an example of either form of AI. Most interesting, and presumably most difficult, is to test for features that are generally considered to be more specifically human and which presumably would be part of AGI. Ray Kurzweil, for example, despite his strong focus on *computational* power, claims that “[c]onsciousness is the most important ontological question.”⁹ Other possibly specifically human abilities which could be part of AGI are creativity, emotions, or free decision-making. The question of whether such abilities can be created in machines can be posed at least from an epistemic or metaphysical point of view. The issue of testing features that have a subjective component or in some sense require a first-person perspective runs into epistemic problems. Quite obviously, it is not *ultimately* possible to decide whether another being has a first-person perspective, emotions, consciousness, or self-consciousness. This problem has its parallel in the well-known problem of “other minds.” For all we know, my neighbor *could* be a mindless zombie, as in Chalmers’ Zombie argument. However, this epistemic problem will be left aside for the moment.

Also, rather than arguing for the possibility of AGI and discussing arguments such as Searle’s Chinese Room Argument, which claim that AGI is not possible,¹⁰ I will *assume* that it is, in principle, possible to create and develop an AGI and, subsequently, even an ASI. This assumption is supported by many researchers such as Nick Bostrom, Ray Kurzweil, and Max Tegmark, who argue for the possibility of AGI, implicitly or explicitly from the point of view of naturalistic or physicalistic frameworks.¹¹ Importantly, assuming the possibility of AGI does not say anything about *how* a system which realizes, for example, mental states may be structured. At least two lines of development can be imagined. Either AGI is realized in a non-biological medium, in silicon, for example, or it is realized by gradually exchanging parts of a human with machine parts; the human first becomes a cyborg and eventually an AGI. Also, it may, for example, turn out to be impossible for humans to construct such systems due to their limited cognitive capacities, or silicon-based systems like our computers may not be suitable for the project of AGI; a project which according to the above would include features like consciousness, self-consciousness or mental states in general. Furthermore, following Lenzen, I believe that at present *none* of the examples of machines’ intelligent *behavior* are examples of AGI or ASI. Deep learning algorithms such as chatbots *are not* implementations of AGI. A chatbot is usually restricted to certain types of conversations; deep learning algorithms perform specific tasks selected by human programmers. AlphaGo, which implements deep learning algorithms and thus plays Go better than any human, does not exhibit conscious behavior in the same way as a human Go master does. The program, for example, does not have the qualitative experience of enjoying or disliking the game. A self-driving car does not experience the phenomenal quality of the landscape through which it drives. In detail, Lenzen has discussed many more examples involving robots, “creative” programs, chatbots, and the like. She concludes that none of these examples – at least at present – should be regarded as examples of AGI in the sense described here.¹² However, if AGI *were* created, then this AGI surely would incorporate all of the abilities given in these examples: it would be conscious of these abilities, it would

be self-conscious, it would have qualitative experience, and it would surpass human abilities by far in many, possibly even all, areas of expertise.

Anthropomorphism and Anthropocentrism in Human Relations to AI

In the following, I shall understand anthropomorphism as the act of ascribing human traits or abilities to non-human entities, while anthropocentrism in its weaker form shall very broadly be understood as the act of placing human beings at the center in some aspects without *necessarily* regarding humans as most important or superior. In a stronger form, anthropocentrism shall be understood as placing humans at the center in the sense of being *most important* in the world or a specific context and being *superior*, at least in some aspects.

Moreover, I shall make the distinction between unjustified and justified anthropomorphism or anthropocentrism. For example, the anthropomorphic attribution of the ability to love another being to, say, a chimpanzee, may initially seem unjustified or misplaced but may turn out to be justified after all, given some exceptionally loving actions displayed by a chimpanzee. Believing humans are the crown of creation seems to be an example of misplaced or unjustified anthropocentrism since humans do not know whether there are other sentient beings in the universe who are superior to us. Still, anthropocentrism could be justified locally in some cases; it is conceivable that there are cases in which we do not have any choice other than placing humans at the center of reasoning or the center of action. Humans should, for example, place themselves at the center of actions for protecting the earth, not necessarily in the sense that they are most important on earth or superior to other beings, that is, in the sense of the above mentioned stronger understanding of anthropocentrism, but in the sense that on this planet *only* humans can at present be at the center of such actions. It should also be noted that anthropomorphism in psychology is generally acknowledged as typical human behavior, particularly in relation to animals¹³ and that anthropocentrism, even in the stronger sense, is often regarded as deeply ingrained in Western civilization, ethics, and theology.¹⁴ As an example, the concepts of well-being, happiness, or eudaimonia in Aristotelian ethics¹⁵ are anthropocentric, at least in the weaker sense. Anthropocentrism in the stronger sense is, I believe, nevertheless implicitly assumed since striving for happiness, for example, presupposes that happiness of *humans* is a “higher” and possibly even the “highest” good. Thus in this example the needs of humans are more or even most important in the world. In general, humans know in which cases their anthropomorphism is unjustified. If I give a chicken a human name and even attribute some human traits to it, I am, nevertheless, well aware that the chicken is not human and may not have the traits I assigned to it in any stronger sense. Likewise, if a human finds his/her robotic lawnmower cute due to the lights shining like little eyes in the darkness, s/he most probably knows very well that the “cute” little lawnmower is nevertheless a robotic machine, in this case an example of ANI. Thus s/he knows that s/he is in some sense superior to this machine, not in the sense of lawn-mowing, of course, but as a thinking and sentient being in this world. The knowledge of being superior in this specific sense means that the human compares herself with the inferior being, which has unjustifiably been anthropomorphized, and in this measurement and comparison places herself at the center of this act and claims superiority. That is, the human is

implicitly acting in a strong anthropocentric manner. To put it another way, unjustified anthropomorphism directed towards ANI leads to strong anthropocentrism.

Even in cases of ANI which involve higher cognitive abilities, this still seems to be the case. Take the example of the Go-playing computer. Humans assign the human ability to play Go to the computer. The Go program has even been described as playing like a god and has made moves that could not be anticipated by experts, which could be interpreted as instances of creative playing.¹⁶ In other words, concerning the specific ability to play Go, the computer is both superior and creative in its moves. Nevertheless, I believe that almost any human would regard herself as superior in many *other* aspects of being. The computer would still most certainly be regarded as “a mindless and emotionless machine.” A similar relation of humans to AI is described by Diane Proudfoot when she observes that denying robots emotions, while *at the same time* anthropomorphizing them, is not uncommon among AI researchers.¹⁷ Furthermore, David Watson observes that AI and even algorithms used in AI have at least historically been described in anthropomorphic terms.¹⁸

In these examples, both the unjustified anthropomorphic move and the anthropocentric move initially seem to be unproblematic. Why would it be problematic to speak of the AlphaGo program *as if* it played like a god? Why is it problematic to regard oneself as superior to the machine in many other aspects, when this quite obviously is the case?

According to Proudfoot, one problem with anthropomorphizing artificial systems is that it “makes it too *easy* to convince us of the intelligence, human-level or otherwise, of a machine” (my emphasis).¹⁹ She subsequently argues that the Turing-test, although the test measures human-like intelligent behavior, avoids the problem of anthropomorphism *because of* the risk of making an incorrect judgment due to anthropomorphism. The judges in the test will be especially careful and observant towards possible misinterpretations.²⁰ Since the test measures human-like behavior and is thus oriented at the intelligence and behavior of human beings, the test is also anthropocentric in the above-defined weaker sense. As mentioned above, according to Elamrani and Yamploskiy, this seems to be the case in almost any test.²¹

Another problem is the following: If the above seemingly unproblematic anthropomorphism in relation to ANI becomes or already has become a common habit, a way of relating to other non-human beings *in general*, then it may indeed become problematic since the backdoor of *always* claiming that the other-being is “merely a machine” is left open. Humans could always claim that they are unjustly anthropomorphizing the supposed AI-system, be it ANI, AGI, or ASI, and that this system is, nevertheless, merely “a mindless machine” and that the machine will *never* achieve any truly human mental abilities. The claim that it is always unjustified to anthropomorphize AI-systems, even if the machine exhibits abilities which could or even should be comparable to human abilities, seems to open up the possibility for a more general strong anthropocentrism.²²

John Danaher discusses anthropomorphism in relation to AI in terms of deception. He distinguishes between three different forms in which a robot can be deceptive: external state deception, superficial state deception, and hidden state deception. Danaher believes that the first case in which “[t]he robot uses a deceptive signal regarding some state of affairs in the world external to the robot”²³ is neither philosophically nor ethically unique. A lie from a robot would be an example of this case, and would presumably have similar ethical consequences as a lie from a human.²⁴ In the second case, the robot is

deceptive regarding capacities or internal states it actually *lacks*, while in the third case, the robot is deceptive regarding capacities or internal states it actually *has*.²⁵ The examples of anthropomorphism in relation to ANI given here are all examples of the second type of deception in Danaher's distinction. An ANI lacks the capacities or internal states in question, and thus anthropomorphism in relation to it would be unjustified in the sense in which I have introduced it here. Danaher believes that it is common to lump the second and third categories together and that this is unfortunate, since the third type of deception is ethically harmful as it can be read as robot *betrayal*.²⁶ While the third form of deception would have to be intentionally implanted into an ANI by its human maker, an AGI or ASI with full consciousness and mental abilities on par or exceeding human abilities would certainly be able to deceive another being about its capacities by itself.

Also, according to Watson, unjustified anthropomorphism in, for example, cases of decision-making algorithms in socially sensitive applications, can lead to the problem that the responsibility for the results and consequences of such applications is handed over to other groups of humans who control these systems.²⁷ In this case, anthropomorphizing, if not recognized as such, places other humans at the center, namely those who have the power of controlling the system and realize that the anthropomorphism is actually unjustified.

A further important observation is the following: At first glance, anthropomorphism towards ANI seems to be unjustified in general. However, a closer look reveals that anthropomorphism is to some extent justified after all. Consider the Go-playing computer again. To anthropomorphize the computer *in total* is surely unjustified; it is still emotionless and mindless. Nevertheless, if one merely focuses on the ability to play Go, then anthropomorphism is justified since the computer not only plays Go *as if* it were superior but, in fact, *is* superior. Thus there may always be some, possibly minor, aspect in which the anthropomorphism is justified, and humans should not in an anthropocentric manner regard themselves superior at least in that minor aspect. However, humans could still claim that any given ability in which an ANI is superior is an *extension* of the human ability, thus still upholding the strong anthropocentric position by choice.

It should be noted here that anthropomorphizing *how* the algorithm works may nevertheless be unjustified. Although, for example, the naming of deep *learning* algorithms suggests a likeness to human learning, such similarities are generally overstated.²⁸ Likewise, it should be emphasized that there are significant differences in the structure of the "neuronal" networks used in deep learning compared to biological neuronal structures.²⁹ Further, deep learning algorithms are flawed in ways human learning is not. It is, for example, well-known that a deep learning algorithm requires vastly more data to recognize a cow, compared with a child who can typically recognize a cow after an adult has *once* shown the child what a cow is.³⁰ However, the focus here is rather on the actual outcome of the actions of an AI. Would the implemented ability be comparable to a human ability, and would it be justified to be anthropomorphic with respect to *what* that ability achieves? In the cases of unjustified anthropomorphism, it seems at least easy to maintain a strong anthropocentric position, while in the cases of justified anthropomorphism there appears to be less motivation to be anthropocentric in the stronger sense.

Before turning to the cases of AGI and ASI, I will provide a brief look at occurrences and problems of anthropocentrism in the context of AI and why strong anthropocentrism, in general, is problematic. It has already been mentioned that research in AI is at least initially anthropocentric. The project of developing systems that perform cognitive abilities is oriented towards *human* abilities.³¹ Also, as mentioned above, testing consciousness is anthropocentric in the weaker sense. Such weak anthropocentrism, I believe, cannot be avoided. Since there are no other beings that can perform tests for consciousness, humans do not have a choice but to perform the tests themselves and take as a measure what they *know* to be conscious, namely their own human consciousness. It is worth mentioning that Max Tegmark, in his study of AI, *Life 3.0*, attempts to define consciousness in terms that avoid any anthropocentric connotation; he simply equates consciousness with subjective experience.³² However, anthropocentrism in the weaker sense is not avoided by this definition since both subjective and experience are terms that humans only understand by precisely referring to *human* subjectivity and experience.

The problem with anthropocentrism is more general, which is possibly one of the reasons why Tegmark, in his definition of consciousness, wishes to avoid it. In particular, it is well-known that anthropocentrism has been criticized for its negative consequences on the human relationship to the world and other beings in general.³³ Thus the above conclusion that there is a danger of ending up in strong anthropocentrism in relation to AI would be a problem due to the general problem of anthropocentrism. Examples of weak anthropocentrism, as in the tests for consciousness, seem to be hard to avoid. However, at least in the case of tests for consciousness, this would presumably be unproblematic since the researcher who is testing for consciousness is in general aware that the test is structured with reference to human consciousness, human intelligence or some other human trait, and that it may be hard to find another suitable reference point. Nevertheless, I believe it is important to highlight in which cases a test is designed with some form of implicit anthropocentrism.

Proceeding to the cases of AGI and ASI, a further question is how these relate to anthropomorphism and strong anthropocentrism? Assume for the moment, as I have suggested earlier, that AGI and even ASI are possible to create and that they even exist. For the sake of simplicity, I shall henceforth only consider AGI, since in all likelihood AGI entails that ASI is also possible. AGI would already be superior to humans in many ways since it would be able to incorporate all of the already existing implementations of specific human abilities present in ANI. Thus in this thought experiment, AGI and ASI can safely be treated in the same manner.

Now imagine that AGI is implemented and realized, that humans encounter another being of their creation, which is intelligent, conscious, self-conscious, and so forth. To regard the assumed mental or cognitive abilities of such a “machine” as on par and even superior to the corresponding human abilities seems to be highly plausible. Thus assigning such abilities to this “machine,” anthropomorphizing seems to be justified even if such mental abilities may not be implemented by methods that can be *described* in anthropomorphic terms. Surely, through the human freedom of choice, it would still be possible to claim that humans are unjustly anthropomorphizing the AGI system and that this system is, after all, merely “a mindless machine.” Humans could simply deny any possible human traits or abilities which parallel human abilities. However, wouldn’t that be to claim that the machine is mindless just because it is a *machine*? Wouldn’t that mean

that one is simply choosing to be anthropocentric? David Gunkel, for example, discusses the alterity of machines in *The Machine Question*. He wishes to avoid the underlying metaphysical and epistemological problems related to the ontological status of mental abilities in AGI and the question of establishing in which cases an AI can be said to possess them.³⁴ With reference to the Information Ethics developed by Luciano Floridi and J. W. Sanders, he argues that although Floridi and Sanders wish to replace anthropocentric or biocentric theories with more ontocentric ones, that one form of centrism is simply replaced with another.³⁵ In conclusion, Gunkel suggests that questions concerning moral agency or moral subjectivity of AI, that at least partly question whether it is justified to regard an AI as having human traits, need to be answered individually on a case by case basis.³⁶ This is surely a possible approach which is also reasonable in practice; however, it does not ultimately answer the above question of whether it is justified to claim that the machine is mindless just because it is a *machine*, nor does it answer whether it is, in general, unjustified to anthropomorphize an AGI-system or not.

Rather, my point is that once I realize and accept that it is *justified* to anthropomorphize a specific ability, it becomes unjustified to be anthropocentric, at least in the stronger sense, in relation to the same ability. That is, in this case I should *not* choose to be anthropocentric. Humans can then not claim to be superior concerning that particular ability. Surely, in the case of ANI, humans could still uphold the strong anthropocentric position, since on the whole it is mindless. One may also wonder whether such justified anthropomorphism is still anthropomorphism; whether the terminology is still suitable. Be that as it may, the realization of a “machine” with the abilities of an AGI, even if it is of human creation, would point towards the fact that humans cannot and should not continue to be anthropocentric in the stronger sense. The mere existence of such a machine would strongly suggest that humans should realize that they are not at the center of the universe anymore. This move away from anthropocentrism would have considerable force; humans would not have an actual choice, as in specific cases of justified anthropomorphism, to be anthropocentric or not. Of course, humans could *refuse* to accept that AGI is on par or even superior to themselves. However, in the long run, I believe they would simply have to realize that they are not at the center anymore. James Lovelace suggests that the creation of AGI would be the beginning of a new era in the history of humanity, namely the end of the Anthropocene and the start of the *Novacene*.³⁷

In summary, unjustified anthropomorphism leads to strong anthropocentrism either because the anthropomorphizing humans realize their anthropomorphism and end up claiming that the machine is inferior in many if not all aspects of mentality, or because they naively stick to the unjustified anthropomorphism they hand over control to other humans who will in turn make the strong anthropocentric claim that humans are superior. Moreover, there is a danger in claiming that anthropomorphism *is always* unjustified, which generally supports strong anthropocentrism; if all human traits in another being or system are denied, humans can easily be regarded as superior, in the sense of strong anthropocentrism. It has further been emphasized that the problem of strong anthropocentrism is of a more general nature. Concerning AGI, it has been argued that anthropomorphism in some sense becomes justified and that consequently, humans should not and possibly even cannot continue to be anthropocentric. The question is whether this anti- or non- anthropocentric conclusion (henceforth I will use non-

anthropocentric to mean “not being anthropocentric in the stronger sense”) leads to dehumanization?

Pantheism, Anthropocentrism, and AGI

In the following section, I argue that a consequence of pantheism is for human relations to the world to be similar to the above conclusion regarding the rejection of anthropocentrism. Moreover, I suggest that pantheism provides a fruitful way for humans to relate to the world in general and AI in particular, and that the rejection of anthropocentrism in pantheism even suggests a further step towards cosmocentrism.

Pantheism can be defined in generic terms as follows: (1) the divine is transcendent, that is the divine in some sense is greater than the world, (2) the divine is immanent, that is the world is in the divine or imbued by the divine, and (3) the world and the divine have a bilateral relationship, that is the divine affects and is affected by the world.³⁸ This very general form of pantheism need not necessarily be understood as a form of theism in the sense of the divine being a personal God; it could indeed be read as a form of pan-en-divinity or cosmopsychism. Yujin Nagasawa and Khai Wager have recently discussed the latter, for example.³⁹ I believe that keeping the definition more general or even adopting the terms pan-en-divinity or cosmopsychism has the important advantage that the conclusions to be drawn here can easily be generalized or transferred to other contexts and therefore are presumably not restricted to a specific form of theism such as Christianity, Islam, or Judaism. Nevertheless, I shall employ the more well-known and common term of pantheism and even use some concepts common to the Christian tradition.

Thus, how does such generic pantheism relate to anthropocentrism? From the definition above it follows that anything, be it the forest outside my window, the air I breathe, my friends, relatives, and workmates, etc., is permeated by the divine. If this thought is fully embraced, then anything created by humans would *also* be permeated by the divine. Thus if I wish to embrace the idea of pantheism truly and deeply, then I should believe that the divine is manifested in *anything* and everything I encounter in this world. Admittedly, this may initially seem to be an odd thought. Is, for example, a rusty old car in some sense divine? However, I believe that an important point is that humans should, and perhaps even need to, find ways to see the divine in anything. Since humans themselves are also in some sense divine by merely being part of the world, then at least attempting to see the divine in anything, regardless of how difficult this may turn out to be, would bridge the gap between the perceiving subject and the perceived object. It would put humans in the midst – not at the center – of nature or the created and co-created world.

Furthermore, a genuinely bilateral relationship to the world, which is in balance, would not imply a hierarchy. Both humans and the rest of creation would be permeated by divinity. In that sense, their divinity would be on par. Undoubtedly, one could argue that there are differences in divinity with respect to a being’s consciousness or its being a moral agent or patient, which would nevertheless imply some form of hierarchy, and which could be read in terms of “more” or “less” divine. The point here is, however, that divinity is everywhere and that divinity, by already being inherently divine cannot, just like infinity, be compared in terms of “more” or “less.”

In the context of feminist and ecological theology and panentheism, Sallie McFague interprets this all-present divinity and argues that humanity needs “to turn away from anthropocentrism – the focus on ourselves as masters of the earth – to cosmocentrism – the focus on the earth and where we belong.”⁴⁰ Similarly, Jürgen Moltmann concludes that any experience may have a transcendent inner and that anthropocentrism is supported by a dualistic view of transcendence and immanence, which in turn should be overcome.⁴¹ Thus, fully accepting strong immanence with its relationality in panentheism would mean the negation of any unjustified or misplaced anthropocentrism in the stronger sense defined above since, due to the divinity of everything, nothing could ultimately be regarded as inferior to humans. At this point, it is important to emphasize again that anthropocentrism can still be justified in some cases, however *not* in the stronger sense. In the above-mentioned case of striving for happiness or well-being, it would be justified to strive for such well-being, however *without* placing human needs at the center or affording them the highest priority. By the relational, non-anthropocentric position suggested here, humans would have to reason about what happiness or well-being is on the grounds of their relationality to the world and thus to the all-present divinity in a panentheistic worldview. Importantly, it is up to humans themselves to realize the divinity in everything and thus cease to be anthropocentric in the stronger sense.

In a Christian context, it would furthermore be possible to apply the doctrine of *imago Dei* to human relations to the world. If humans, on the one hand, are created in the image of God, but on the other hand, as Herzfeld has suggested, create intelligent machines in the image of humans, in *imago hominis*, then the machines are in some sense at least indirectly also created in the image of God. Thus ANI, AGI, or ASI would all be created in the image of humans and consequently in the image of God. Mental abilities like consciousness or self-consciousness would not only be reflections of the human abilities but also, although indirectly, reflections of divinity. To be sure, the divinity of such abilities is already established in panentheism by God’s immanence. However, the doctrine of *imago Dei* and Herzfeld’s suggestion that one of the motives for creating AI is to create beings in the image of humans further underlines this divinity.

Interestingly, Herzfeld believes that one of the reasons that AI is regarded as far less problematic in Japanese society than in Western societies lies in the *animistic* connotations of Shintoism.⁴² If she is correct in her observation, then such animism could be interpreted analogously to the pan-en-divinity suggested here, thus providing further support for the thesis that mental abilities can, in general, be seen as expressions of the divine.

Even in the example of an ANI, in the Go-playing machine, the *ability to play Go* is, in a broader sense, imbued by divinity. In an anthropomorphic view, the machine is regarded by the human player as an opponent, who considers strategies, moves, and tactics, and can possibly win the game. As explicated above, anthropomorphism is justified in relation to the ability as such, however not necessarily in relation to ascribing, for example, strategic considerations or tactics to the machine. The above reasoning about panentheism, *imago Dei*, and *imago hominis* would support the view that anthropomorphism in relation to the ability as such is justified, since the ability is a reflection of the *human* ability to play the game, which in turn is a reflection of the divine in total. Thus, firstly, through the above conclusion about the relation between anthropomorphism and strong anthropocentrism, the latter would be unjustified in relation to this

specific ability. Secondly, following the general reasoning above, this conclusion could be generalized and extended to the case of AGI. Seeing the so-called artificial abilities – it may be preferable to use the term created abilities – of an AGI as a reflection of human abilities seems to be justified, and if it is thus justified to view these abilities in anthropomorphic terms, then through the above reasoning it is not justified to be anthropocentric in relation to these abilities. Finally, panentheism would furthermore suggest that such abilities, both in ANI and AGI, should in some sense be regarded as permeated by divinity.

If the created artificial abilities in AGI and ANI are regarded as reflections of human abilities permeated by the divine, the border between what is human and artificial, or “natural” and artificial, seems to be erased. This can be interpreted as a call to overcome dualistic distinctions such as human/artificial or natural/artificial. Instead of placing humans at the center, humans should realize their position *between*, in the *midst* of creation, including the “artifacts” created by them. “Being in the midst” or “between” does not necessarily result in a hierarchical view or the view that either side should rule over the other. Thus, as a consequence, interpretations of Biblical passages in terms of stewardship rather than dominion over the world should be preferable.⁴³ Since this consequence includes ANI, AGI, and ASI, once again, the question arises whether such anti- or non-anthropocentrism leads to dehumanization and whether rejecting anthropocentrism conflicts with humans striving for their well-being? Should humans regard themselves as “lesser” beings given the proposed existence of AGI?

Initially, this may appear to be the case. A supposed AGI would most probably have superior abilities in any mental or cognitive aspect, and presumably even in other aspects. If it also, by the assumption of being an AGI, exhibits consciousness, self-consciousness, and creativity, for example, would it not then be superior? Surely in some sense, this would be the case. However, is realizing that something or someone is superior dehumanizing? Even without AGI’s proposed existence, there will always be other humans or other beings, which, in some or many aspects, are superior. That is certainly not dehumanizing. Here thinking in terms of panentheism is again helpful. Firstly, humans also are permeated by the divine. Secondly, if humans would become “lesser” in light of AGI’s possible creation, then in contradiction to the above conclusion that humans stand in the midst of creation, humans would *not* stand in between or in the midst of creation.

Moreover, if our being human would be *defined* merely by *strong* anthropocentrism, if one of the defining traits for being human would be to be superior and at the center of creation, then overcoming this merely strong anthropocentrism would be dehumanizing. However, as hinted above, humans would still have to reason about what human well-being could be on the grounds of *relations* to the world, which does not presuppose that humans are at the center or have the highest priority. Further, does the suggestion of overcoming *strong* anthropocentrism, as argued for both in a non-theological setting in relation to AGI and in parallel in a panentheistic setting, not actually point at what it means to be human? Indeed, the pan-en-divinity in panentheism points even further – it points to the rejection of strong anthropocentrism *and* cosmocentrism. Moreover, in the case of the argument for non-anthropocentrism in the non-theological setting, humans may not have a choice other than to accept non-anthropocentrism in the face of a fully developed AGI. In contrast, in the case of pan-en-divinity, humans still have the choice of realizing the divinity in everything and can consequently choose to be non-

anthropocentric. The panentheistic suggestion of standing in between or in the midst of creation instead, seems to invite humans to live in *harmony* with both what has been created – nature, for example – and what is created by humans – AI, for example. Importantly, it should be noted that standing in between does not imply that humans are machines or vice versa, an implication which Herzfeld considers to be dangerous.⁴⁴ It instead suggests that humans should establish a non-hierarchical relationship to fellow humans, creation in general, and AI in particular; a suggestion supported by Herzfeld, who also emphasizes the development of a positive human-AI relationship.⁴⁵

Does this more positive attitude towards AI in general and the possible development of AGI mean that there are no dangers with AI technology? Certainly not. Applying the doctrine of *imago Dei* or Herzfeld's notion of *imago hominis* in a panentheistic setting to AI does not say anything about *how* human abilities are reflected in AI. It may well be the case that humans choose to reflect the more negative sides of their abilities in AI. Indeed, I believe that dystopic depictions of AI in popular culture partly reflect the fear of how humans would react if they had the upper hand; they reflect the darker sides of ourselves. However, standing in between, in the midst of creation with a more positive attitude towards AI rather hands back the responsibility to humans. Instead of engaging in the strong anthropocentric fear that something, some AI perhaps, may become *superior* to humans, humans should in a weak anthropocentric sense realize their responsibility for what should and will be implemented in future AI-systems; they should also realize their role and position in creation and towards their creation. Thus, taking seriously the non-anthropocentric move as argued for in a non-theological setting and in parallel in a panentheistic setting, would call for humans to seek what it is to be truly human, to “know thyself” – γνῶθι σεαυτόν, as inscribed in the Temple of Apollo at Delphi. Surely it is not to be truly human to strive for superiority and dominion while at the same time fearing the loss of superiority and dominion in light of the possible *self-initiated* creation of a superior being. Rather, one possible way to “know thyself” would be to practice human freedom and *freely* realize the role of humans as caretakers of creation,⁴⁶ to see *all creation*, both human and divine, as ultimately divine, and to accept ourselves as part of creation rather than the center of creation. This stands in contrast to the non-theological line of reasoning in which humans would most probably be coerced into the direction of giving up anthropocentrism if and as soon as AGI is realized, while in the case of panentheism humans would have the *freedom* to choose the non-anthropocentric stance and see the possible benefits of such a view both in relation to the world in general and in relation to the development of AI.

Summary

Under the assumption that AGI is possible, and that humans should expect AGI to be developed based on structures other than the biological structures found in human brains, it has been argued that the more it is justified to ascribe human abilities to AI-systems, the less it is justified to be anthropocentric in relation to the same abilities. Thus, an AGI with abilities equal or superior to human abilities would point to the view that humans should or can no longer be anthropocentric in the stronger sense. Similarly, the reasoning about the relation of panentheism to anthropocentrism in general

and AI in particular led to the conclusion that humans, by their own choice, should be non-anthropocentric and even cosmocentric concerning their position in creation. They are in the midst of creation, not at the center. It has also been argued that the non-anthropocentric move is not dehumanizing; it instead invites humans to seek for what is truly human.

Undoubtedly, many problems and questions regarding human relations to AI have not been discussed here. One important problem is the legitimacy of results of AI-systems such as AI-based decision-making in liberal societies, which raises questions about epistocracy or algocracy.⁴⁷ Another example would be questions regarding the use of social robots in childcare, geriatric care, or care in general. However, realizing and implementing the call for a non-anthropocentric approach and attitude to the world in general and AI in particular, as has been directly argued for here by the imagined development of AGI and by the implications of panentheism, may turn out to be fruitful in further reflections on such topics.

Notes

1. John R. Searle, *Mind: A Brief Introduction* (Oxford: Oxford University Press, 2004), 46.
2. David J. Chalmers, *The Conscious Mind* (Oxford: Oxford University Press, 1996), 313.
3. Manuela Lenzen, *Künstliche Intelligenz* (München: C.H. Beck, 2018), 31–34.
4. Ted Peters, “Intelligence? Not Artificial, but the Real Thing!,” *Theology and Science* 17:1 (2019), 3.
5. Noreen Herzfeld, *In Our Image* (Minneapolis: Fortress Press, 2002), 33.
6. John McCarthy, “From Here to Human-Level AI,” *Artificial Intelligence* 171:18 (2007), 1174–1182; Andreas Kaplan and Michael Haenlein, “Siri, Siri, in My Hand: Who’s the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence,” *Business Horizons* 62:1 (2019), 16.
7. Andreas Kaplan and Michael Haenlein, “Siri, Siri, in My Hand: Who’s the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence,” *Business Horizons* 62:1 (2019), 18–19.
8. Aïda Elamrani and Roman V. Yampolskiy, “Reviewing Tests for Machine Consciousness,” *Journal of Consciousness Studies* 26:5–6 (2019), 50.
9. Ray Kurzweil, *The Singularity Is Near* (Richmond: Duckworth Overlook, 2005), 380; Yujin Nagasawa, “Panpsychism Versus Pantheism, Polytheism, and Cosmopsychism,” in *The Routledge Handbook of Panpsychism*, ed. William Seager (New York: Routledge, 2019); Yujin Nagasawa and Khai Wager, “Panpsychism and Priority Cosmopsychism,” in *Panpsychism*, ed. Godehard Brüntrup and Ludwig Jaskolla (Oxford: Oxford University Press, 2017), 113–129.
10. For example, Searle, *Mind: A Brief Introduction*, 90.
11. Nick Bostrom, *Superintelligence* (Oxford: Oxford University Press, 2014); Ray Kurzweil, *How to Create a Mind* (Richmond: Duckworth Overlook, 2012); Max Tegmark, *Life 3.0* (New York: Vintage Books, 2017).
12. Lenzen, *Künstliche Intelligenz*.
13. Kristin Andrews, “Animal Cognition,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N Zalta, Summer 2001 (Metaphysics Research Lab, Stanford University, 2016). Andrews describes anthropomorphism as either a category mistake or a false attribution. Nicholas Epley discusses anthropomorphism from the point of view of psychology and consumer research. Nicholas Epley, “A Mind Like Mine: The Exceptionally Ordinary Underpinnings of Anthropomorphism,” *Journal of the Association for Consumer Research* 3:4 (2018), 591–598.

14. See, for example, Andrew Brennan and Yeuk-Sze Lo, “Environmental Ethics,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N Zalta, Winter 2001 (Stanford: Metaphysics Research Lab, Stanford University, 2016); Beth Preston, “AI, Anthropocentrism, and the Evolution of ‘Intelligence’,” *Minds and Machines* 1:3 (1991), 274; Sallie McFague, *A New Climate for Theology* (Minneapolis: Fortress Press, 2008), 46–49.
15. Aristotle, *Nicomachean Ethics*, trans. H. Rackham, Loeb Class (Harvard: Harvard University Press, 1926).
16. For example, Alex Walker, “World’s Best Go Player Describes Google’s AI As A God,” 2017, <https://www.kotaku.com.au/2017/05/worlds-best-go-player-describes-googles-ai-as-a-god/>.
17. Diane Proudfoot, “Anthropomorphism and AI: Turing’s Much Misunderstood Imitation Game,” *Artificial Intelligence* 175:5–6 (2011), 951.
18. David Watson, “The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence,” *Minds and Machines* 29:3 (2019), 417–440.
19. Proudfoot, “Anthropomorphism and AI,” 951.
20. *Ibid.*, 252–253.
21. Elamrani and Yampolskiy, “Reviewing Tests for Machine Consciousness,” 50.
22. This rejection of anthropomorphism in general amounts to the denial briefly discussed by Arleen Salles, Kathinka Evers, and Michele Farisco. In contrast to this article, they conclude that “Understanding AI through the lens of human mental features risks reducing it to a sort of replica of the human mind and leads to a flawed and ultimately limited ethical analysis of the issues AI raises.” Arleen Salles, Kathinka Evers, and Michele Farisco, “Anthropomorphism in AI,” *AJOB Neuroscience* 11:2 (2020), 88–95. <https://doi.org/10.1080/21507740.2020.1740350>. However, I believe that the conclusion in this article does not lead to the consequence that similarities between AI and humans are overemphasized. An implicit goal of this article is to find a position in which humans can live in harmony with both themselves and what they create.
23. John Danaher, “Robot Betrayal: A Guide to the Ethics of Robotic Deception,” *Ethics and Information Technology* 22:2 (2020), 117–128.
24. *Ibid.*, 121.
25. *Ibid.*
26. *Ibid.*
27. Watson, “The Rhetoric and Reality of Anthropomorphism, 417–440.
28. *Ibid.*, 417–440.
29. Salles, Evers, and Farisco, “Anthropomorphism in AI,” 92.
30. Watson, “The Rhetoric and Reality of Anthropomorphism,” 422–423.
31. Preston, “AI, Anthropocentrism, and the Evolution of ‘Intelligence’,” 262.
32. Tegmark, *Life 3.0*, 38, 283.
33. See for example in the context of AI, Victoria Davion, “Anthropocentrism, Artificial Intelligence, and Moral Network Theory: An Ecofeminist Perspective,” *Environmental Values* 11:2 (2002), 163–176; Preston, “AI, Anthropocentrism, and the Evolution of ‘Intelligence’.”
34. David Gunkel, *The Machine Question* (Cambridge, MA: The MIT Press, 2012).
35. *Ibid.*, 151.
36. *Ibid.*, 215.
37. James Lovelock, *Novacene* (Colchester: Penguin Books, Allen Lane, 2019).
38. Niels-Henrik Gregersen, “Three Varieties of Panentheism,” in *In Whom We Live and Move and Have Our Being*, ed. Philip Clayton and Arthur Peacocke (Cambridge: Wm. B. Eerdmans Publishing Co., 2004), 22; Mikael Stenmark, “Panentheism and Its Neighbors,” *International Journal for Philosophy of Religion* 85:1 (2019), 23–41.
39. Nagasawa and Wager, “Panpsychism and Priority Cosmopsychism,” 113–129; Nagasawa, “Panpsychism versus Pantheism, Polytheism, and Cosmopsychism.”
40. Sallie McFague, *The Body of God* (Minneapolis: Fortress Press, 1993), 144; McFague, *A New Climate for Theology*, 49.
41. Jürgen Moltmann, *Der Geist Des Lebens* (Gütersloh: Gütersloher Verlagshaus, 1991), 44–51.

42. Noreen Herzfeld, *Technology and Religion* (Conshohocken: Templeton Press, 2009), 130. While I believe that Herzfeld's claim is plausible, a more thorough investigation and study would be valuable to support her claim. From personal experience, it seems that Japanese people have a more positive attitude even towards the implementation of AI in social interaction. In a seminar held at Linköping University, in September 2019, a Japanese speaker argued for the use of robots in childcare, for example.
43. McFague, *A New Climate for Theology*, 64–66; Ian G. Barbour, *Nature, Human Nature and God* (Minneapolis: Fortress Press, 2002), 124–125.
44. Herzfeld, *In Our Image*, 88–94.
45. *Ibid.*, 93.
46. See, for example, Gen 2:15.
47. John Danaher, "The Threat of Algocracy: Reality, Resistance and Accommodation," *Philosophy and Technology* 29:3 (2016), 245–268.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes on Contributor

Dr. *Oliver Li* is a researcher at the Center for Multidisciplinary Research on Religion and Society (CRS Uppsala) in the Department of Theology at Uppsala University, Sweden. He is a project member of the multidisciplinary projects Artificial Intelligence, Democracy and Human Dignity and The Artificial Public Servant funded by the Marianne and Marcus Wallenberg Foundation.