Imago Dei in the Age of Artificial Intelligence: Challenges and Opportunities for a Science-Engaged Theology

Marius Dorobantu

Abstract: Modern developments in evolutionary and cognitive science have increasingly challenged the view that humans are distinctive creatures. In theological anthropology, this view is germane to the doctrine of the image of God. To address these challenges, *imago Dei* theology has shifted from substantial toward functional and relational interpretations: the image of God is manifested in our divine mandate to rule the world, or in the unique personal relationships we have with God and with each other. If computers ever attain human-level Artificial Intelligence, such *imago Dei* interpretations could be seriously contested. This article reviews the recent shifts in theological anthropology and reflects theologically on the questions raised by the potential scenario of human-level AI. It argues that a positive outcome of this interdisciplinary dialogue is possible: theological anthropology has much to gain from engaging with AI. Comparing ourselves to intelligent machines, far from endangering our uniqueness, might instead lead to a better understanding of what makes humans genuinely distinctive and in the image of God.

Keywords: artificial intelligence; human distinctiveness; *imago Dei*; relationship; vulnerability

---

Marius Dorobantu is a Theology & Science lecturer at the Vrije Universiteit Amsterdam and a fellow of the ISSR. His doctoral dissertation (at the University of Strasbourg) investigated the potential challenges of human-level AI for the theological understanding of human distinctiveness and the image of God. The article presented here was supported by the Templeton World Charity Foundation under Grant TWCF0542. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Templeton World Charity Foundation.

In 2016, AlphaGo, a computer program developed by Google DeepMind, defeated one of the greatest human players of all time in the ancient strategy game of Go. For many, this event might not have been too significant. After all, computers had already mastered the much more popular game of chess for two decades, ever since Gary Kasparov's famous 1997 defeat by IBM's program, Deep Blue. For me, however, the news about AlphaGo was shattering. Having been an avid practitioner of the game for the best part of my life—both competitively and recreationally—I had a very good idea why this achievement was much more significant than Deep Blue's.

Originating more than four thousand years ago in China, the game of Go has deceptively simple rules. Two players, black and white, compete for limited resources by alternatively placing identical round pieces on a square board, trying to surround more territory than the opponent. Nevertheless, despite the simplicity of the rules, the ensuing complexity of the battle for territory dwarfs any other game. With each move, new possibilities open up, resulting in a cascading number of choices. There are more possible Go games than atoms in the observable universe.[1] For a long time, this made Go inaccessible to computers because the methods used to master other games, such as chess, were simply inapplicable to Go.

Traditionally, computers defeated human players in strategy games by leveraging their superior computing capabilities. Suppose a computer can go through all the relevant possible combinations of a situation on the board in a reasonable amount of time. In that case, there is no need for it to *understand* the game's principles or come up with clever strategies. It simply calculates all the possibilities and selects the one that most probably leads it to victory. In informatics terms, this is called *brute force,* and it is through brute force that Deep

1 David Silver and Demis Hassabis, "AlphaGo: Mastering the Ancient Game of Go with Machine Learning," *Google AI Blog* (blog), 2016, http://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html.

Blue won against Kasparov.[2] In other words, a computer does not need to be clever if it can just laboriously explore all the possible routes. Due to its gargantuan complexity, Go does not lend itself to brute-force calculation. For this reason, the general feeling in the tech community was that it would take at least a few more decades until computers became capable of playing Go at a human level. Hence my surprise!

Besides the computational dimension, there was something more about AlphaGo's achievement that prompted the theologian in me to take notice, having to do with a more mystical aspect of the game of Go. When Go masters explain their moves, they rarely talk in mathematical terms. To be sure, their calculation abilities are outstanding and instrumental for success in the game. But Go masters often revert to a different kind of language when describing their play, one that belongs to the aesthetic register: *it felt good* to play there, or that move *looked beautiful*. A true Go master does not simply look to gain more points than the opponent; she looks for harmony on the board in a way not too different from a painter trying to achieve harmony on a canvas or a musician composing a masterpiece. Therefore, it is unsurprising that the game of Go was included among the four essential arts in ancient China, alongside music, calligraphy, and painting. There is as much calculation involved in a human master's game as intuition, creativity, and aesthetic taste. Moreover, there is arguably also a moral dimension to the game, at least when played by humans. A successful tactic presupposes an ideal mix of character virtues such as patience, humility, courage, and temperance. On the contrary, greed, arrogance, timidity, or pettiness are usually detrimental.

All the above are very subtle and elusive capacities that sit at the core of what we think it means to be human. It is hardly surprising that computers can beat us at chess by simply calculating the most relevant developments in advance. But if computers can beat us at Go, some hard questions arise about what they might become capable of in the

2 Paul Harmon, "AI Plays Games," *Forbes*, 2019, https://www.forbes.com/sites/cognitiveworld/2019/02/24/ai-plays-games/.

future and whether humans and computers are even that fundamentally different.

This article reflects on how progress in AI might impact the understanding of human distinctiveness in Christian theological anthropology, traditionally encapsulated in the notion that humans are created in the image of God (Latin, *imago Dei*). My central thesis is that theological discourse can benefit from engaging with the possibility of human-level AI, despite the apparent devastating impact such a scenario might exert on the idea of human distinctiveness. The analysis begins with a review of current *imago Dei* theology, demonstrating how theological discourse has hugely benefitted from engaging with evolutionary science. The following two sections reflect on how the two main modern interpretations of the divine image might deal with the emergence of intelligent robots. At this juncture, a question will be addressed: could AI be an equally good or even better image of God? The analysis concludes by stating that functional and relational *imago Dei* interpretations could still account for human distinctiveness from intelligent machines, but only insofar as they emphasise the importance of spiritual priesthood, authentic personal relationality, and vulnerability as fundamental human features, instead of rationality and intellectual prowess.

This conclusion demonstrates that theology can benefit from an honest engagement with AI and cognitive science, similarly to how it did by engaging with evolutionary science. Technological developments can bring beneficial limitations for theological speculation by rendering some hypotheses more plausible than others. In other words, it *is* possible for theologians to refine their understanding of human nature and distinctiveness by looking at the kind of intelligences that computer scientists are trying to build. This observation can strengthen the plea for a science-engaged theology. Furthermore, such conclusions regarding what it *really* means theologically to be human can constitute valuable contributions to the interdisciplinary debate on the future of technology. It is still unclear what truly constitutes the marker of humanness, or where does the threshold of personhood lie.

How we answer such questions as a global society will likely have significant ethical implications for how we treat each other, non-human animals, and robots. Theological anthropology can and should, therefore, bring its contribution to this all-important debate.

The image of God after Darwin: Are We Still Special?

"What are human beings, that you are mindful of them?"[3] Since the age of the Psalmist, we have repeatedly asked this question with various methodologies: from theology and philosophy to biology, psychology, anthropology, and cognitive science. So far, none of these intellectual frameworks has come up with complete or satisfying answers.

From the perspective of evolutionary science, we are just one kind of living organism among many others, preoccupied, like all the others, with maximising its survival and procreation while inhabiting a rocky planet that orbits a typical star, just one of the hundreds of billions in the Milky Way. Biologically, we are essentially just another social ape with a slightly larger brain. What distinguishes us from all the other creatures is the things we can do, from writing poetry to sending people to the Moon or contemplating our death. However, all these impressive feats are made possible by anatomical structures and cognitive capacities we share with other creatures, even if they share those capacities in merely rudimentary form: nervous systems, language, mental representations etc. The point is that we do not seem to be as special as we thought we were.

This raises some problems for Christian anthropology because its central tenet is that humans *are* special. After all, they are created "in the image and likeness of God."[4] Since biblical times, we have had this intuition that there must be something special about us, something that distinguishes us from the rest of creation and makes us like our creator. The book of Genesis does not specify what exactly *imago Dei* is, but most interpreters thought of it in terms of some uniquely

3 Psalm 8:4.
4 Genesis 1:26.

human capacity having to do with our intellect, likely influenced by the Aristotelian tradition that regarded humans as rational animals.[5] This is known as the substantive interpretation of *imago Dei*. Nowadays, this interpretation has few adherents because most of the cognitive capacities thought uniquely human in the prescientific age have recently been fully or partially identified in other animals. Furthermore, since Darwin, it has become clear that humans are not ontologically different from the rest of living creatures. We are part of the same tree of life and share most of our DNA—up to 99%—with non-human species.[6]

What does it mean then to be in the image of God, if not to possess some exceptional intellectual faculty? To replace the problematic substantive interpretation, theologians have creatively devised more sophisticated accounts of human distinctiveness and *imago Dei*, most of which broadly fall into two big categories: functional and relational. The functional interpretation locates our specialness not in our mental capacity, but in our election by God,[7] and in what we are called to do, namely, to represent God in the world by exercising dominion and stewardship over the rest of creation. This idea is rooted in the modern biblical exegesis of the notion of image. The assumption is that the image in Genesis was used with a meaning inspired from other cultures in the ancient Near East. To be the image of a particular god, typical of kings or pharaohs, was to represent that god on earth and exercise authority on that god's behalf.[8]

---

5 For reviews of *imago Dei* interpretations, see Noreen L Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit* (Minneapolis: Fortress Press, 2002); Marc Cortez, *Theological Anthropology: A Guide for the Perplexed* (A&C Black, 2010); J. Wentzel Van Huyssteen, *Alone in the World? Human Uniqueness in Science and Theology* (Grand Rapids, MI and Cambridge: William B. Eerdmans, 2006); Stanley J Grenz, *The Social God and the Relational Self: A Trinitarian Theology of the Imago Dei* (Louisville, KY: Westminster John Knox Press, 2001).

6 Robert H. Waterson et al., "Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome," *Nature* 437:7055 (2005): 69–87, https://doi.org/10.1038/nature04072.

7 Joshua M. Moritz, "Evolution, the End of Human Uniqueness, and the Election of the Imago Dei," *Theology and Science* 9:3 (2011): 307–339, https://doi.org/10.1080/14746700.2011.587665.

8 Claus Westermann, *Genesis: An Introduction* (Fortress Press, 1992), 36–37; David J. A. Clines, "The Image of God in Man," *Tyndale Bulletin* 19 (1968): 93.

The other option, the relational interpretation, regards the image of God as manifested in the unique relationship humans are called to have with God and in the authentic personal relationships they have with each other.[9] God, the Holy Trinity, is relationship, and so is humanity because "in the image of God he created them, male and female he created them."[10]

Both these interpretations of *imago Dei* provide better answers to the scientific challenges mentioned earlier than the substantive interpretation. Human distinctiveness does not reside in any uniquely human intellectual faculty but in our unparalleled agency in the world, which we are called to care for and even co-create with God (functional interpretation), or in the relationality that is so central to what it means to be human, and in which we mirror a Trinitarian God (relational interpretation). Although, indeed, we are not the only species that significantly acts upon its environment—many animals, for example, engage in what is known as niche-construction[11]—the sheer scale of our dominion over the earth, at least since the agricultural revolution onwards, might be seen as a proof of our special vocation. Similarly, although we are not the only relational species, the complexity of our personal relationships and the importance of relationships in the development and flourishing of the human person seem to support the idea that it is through our relationality that we are special and in the image of God.

The functional and relational interpretations of the image arguably represent progress from the earlier substantive proposal. This shows that theological anthropology ultimately stands to gain from an open and honest engagement with science. As English theologian Aubrey Moore aptly put it more than a century ago, "Darwinism appeared, and, under the guise of a foe, did the work of a friend."[12] Rev-

9 Karl Barth, *Church Dogmatics*, vol. 3 (Edinburgh: T&T Clark, 1958).

10 Genesis 1:26.

11 Michael Burdett, "Niche Construction and the Functional Model of the Image of God," *Philosophy, Theology and the Sciences (PTSc)* 7:2 (2020): 158–180, https://doi.org/10.1628/ptsc-2020-0015.

12 Francisco J. Ayala, *Darwin's Gift to Science and Religion* (Joseph Henry Press, 2007), 159.

olutionary scientific ideas, such as Copernicus' heliocentric theory or Darwin's evolutionary theory, may appear at first to menace long-held religious beliefs about the world and the human being. Still, once the dust settles, theological reflection is actually enriched by the process of incorporating new scientific knowledge. As it turns out, it is still perfectly possible to speak of a creator God even when we know the cosmos is way older than a few thousand years. Likewise, there are new and arguably better theological ways of speaking of human distinctiveness, even when evolutionary theory shows that we are of the same ilk as nonhuman creatures, and that our cognitive abilities are not that much different in kind from theirs.

However, a new type of challenge for human distinctiveness looms large on the horizon, as hinted at earlier in the AlphaGo story. Starting with the 1950s, computer programs have become capable of matching and surpassing human abilities in an increasing range of tasks, which, when done by humans, require what we vaguely call intelligence. We call this type of program Artificial Intelligence (AI). Even if AI operates somewhat differently from biological intelligence, AI programs are astonishingly capable of doing many of the things we used to regard as the unique domain of human intelligence, such as solving problems, proving theorems, labelling the content of images, transforming speech into text, translating various languages, composing music, and answering questions, to name just a few.

If progress in AI continues, it is not entirely absurd to imagine a time in the future when computers will reach human-level intelligence, becoming able to do all the things that we do equally well or even better. To a certain extent, this is already happening in some domains. AI algorithms can diagnose some forms of cancer better than human doctors.[13] They operate at a superhuman level in chess, Go, and many other strategy games. We trust AI programs to land planes and run the stock markets because of their ability to make fast decisions

13 Scott Mayer McKinney et al., "International Evaluation of an AI System for Breast Cancer Screening," *Nature* 577:7788 (2020): 89–94, https://doi.org/10.1038/s41586-019-1799-6.

better than error-prone humans. One day, our streets might be filled with the much-hyped autonomous cars, or we might engage in deep spiritual conversations with our robotic companions.

When thinking about the challenges posed by AI to the idea of human distinctiveness, the hypothetical scenario of human-level AI is undoubtedly of great relevance. Nonetheless, an argument can be made more broadly that even without such spectacular developments, AI is still relevant for theological anthropology. Here, I would like to refer to AI as more than just the intelligent machines themselves. Instead, AI designates the fundamental study of the nature of intelligence performed by trying to endow machines with intelligence. This is precisely how the field of AI set off in the 1950s. Alan Turing, one of the founders of theoretical computer science and AI, believed that trying to create a thinking machine could shed light on how humans think.[14] In this respect, AI can be seen as an applied form of cognitive science,[15] and its results can be interpreted as saying something relevant about how humans achieve cognition. If AI easily masters chess, Go, prose, or visual arts, this can produce meaningful clues about the nature of such endeavours. On the contrary, if AI stumbles at particular tasks, that is also relevant, perhaps pointing to features that pertain to human distinctiveness. Therefore, both through its successes *and* failures, AI can produce new data points, which can further serve as food for insightful theological reflection.

Could Robots Be Better Images of God?

If AI does reach human level performance, that is, if it matches our ability to *do* things, then the functional interpretation of the image of God as human distinctiveness may become problematic. As long as

14 Jack Copeland, *Artificial Intelligence: A Philosophical Introduction* (Blackwell, 1993), 26.

15 Trying to endow computers with intelligence is one approach. Another approach is the attempt to simulate on supercomputers the neural connections in the mammalian brain: Nidhi Subbaraman, "Artificial Connections," *Communications of the ACM* 56:4 (2013): 15–17, https://doi.org/10.1145/2436256.2436261.

we remain the most capable creature on earth in terms of the things we do, we can still see this as marking our distinctiveness and kinship with God. But how about a scenario where we became stripped of this privileged position by our artificial "mind children"?[16] What if robots became better than humans at ruling the world and, thus, better representatives of God? Should they not, then, also be considered in the image of God (perhaps even more than us?) according to the functional interpretation?

The above hypothesis might look like the stuff of sci-fi movies, but many people in AI take it seriously. In a 2014 survey, 550 AI experts were asked to predict the likelihood of AI reaching the human level soon. The 2040s got a 50% median probability, while the year 2075 got a 90% probability.[17] There is no way of knowing how AI development will continue. Maybe it will slow down and plateau, never really getting anywhere close to the human level. But there is also the opposite scenario, known as the "intelligence explosion,"[18] where progress in AI accelerates, maybe due to machines becoming better than humans at programming AI, thus triggering a positive feedback loop of self-improvement. This scenario is also referred to as the technological "singularity."[19] According to philosopher Nick Bostrom, there is a real possibility that AI could reach a super-human level sometime in the future, something he calls artificial super-intelligence (ASI).[20] We, humans, are severely limited regarding how intelligent we can become. The amount of knowledge we can acquire in a lifetime is limited, our brains cannot grow bigger than our skulls, and they inevitably decay and die after

16 Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence* (Harvard University Press, 1988).

17 Vincent C. Müller and Nick Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," in *Fundamental Issues of Artificial Intelligence*, ed. Vincent C. Müller (Cham: Springer International Publishing, 2016), 555–572, https://doi.org/10.1007/978-3-319-26485-1_33.

18 Ronald Cole-Turner, "The Singularity and the Rapture: Transhumanist and Popular Christian Views of the Future," *Zygon* 47:4 (2012): 787, https://doi.org/10.1111/j.1467-9744.2012.01293.x.

19 Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Penguin, 2005).

20 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, 2014).

several decades. Machines do not share such limitations, and so, in principle, ASI could become more intelligent than any human being, than all of humanity collectively, and even intelligent beyond human comprehension. Bostrom demonstrates quite convincingly that any attempt on our part to contain and control ASI would ultimately be futile because such a super-intelligent agent could see straight through our plans and anticipate any potential strategy we might devise.

There are legitimate concerns about the existential risk posed to our species by ASI, but there are also formidable things that ASI could do for us. The ascension of artificial minds may not happen through a violent rebellion, as often depicted in futuristic movies, but rather with our blessing and cooperation. As our world becomes more complex and data-driven, we will rely increasingly on artificial systems to assist us in our decisions or even to make them in our stead. I mentioned earlier the example of stock markets, which are run by such AI programs, but many other aspects of our lives are already governed mainly by algorithms: what we see in our social media feeds, the music and movies recommended to us by streaming services, how much money we can borrow from a bank, or which medical procedure to choose based on our profile. We are becoming increasingly aware of all the ethical problems associated with this, but it does not seem that we have any intention to reverse this trend anytime soon. Although the loss of privacy and decision-power bothers us in principle, the convenience facilitated by these apps is often too appealing. This is precisely why it is not hard to imagine a future when most, if not all, power is voluntarily granted to AI systems, especially if their competence keeps improving.

Bostrom speaks of three ways ASI might operate: as an oracle, a genie, and a sovereign. As an oracle, it would answer all our questions; as a genie, it would execute all our commands; as a sovereign, it would govern the world with "an open-ended mandate to operate [...] in pursuit of broad and possibly very long-range objectives."[21] Those with a trained theological eye might notice an eerie resemblance to the kind of role ascribed to God in monotheistic religions. But leaving the issue

21 Bostrom, *Superintelligence*, 181.

of idolatry aside, the possibility of ASI governing the world better than we do seems deeply problematic for the functional interpretation of *imago Dei*. How could we still claim to be exceptional if AI proves to be a better steward of creation?

The task is not even that hard to fulfil, given how disastrously we have been performing so far. In our exploitation of animals, we have caused tremendous suffering, especially in the last few decades, with industrial farming. In our greed, we are currently driving the atmosphere to heat up, endangering the ecological balance on a global scale. These *achievements* are hardly something worthy of the divine mandate to represent God in the world. ASI could do a better job, at least in theory. And while that might be something to hope for, from a theological perspective it raises some hard questions about human distinctiveness and our role as stewards of creation appointed through divine election. How could we still speak of such things in a scenario of more-competent-than-humans AI?

I think the question is legitimate, but I do not think a scenario of human-level AI completely invalidates a functional understanding of the image of God. The reason has to do with the scope of our divine mandate to rule over the world, at least as it is understood in many Christian traditions. Our vocation to care for creation goes beyond the historical realm and ultimately has a spiritual dimension. The Romanian-Orthodox theologian Dumitru Stăniloae speaks of a priestly vocation that we are called to, one that enables and compels us to raise the world to a "supreme level of spiritualisation":

> The world was created in order that humanity, with the aid of the supreme spirit, might raise the world up to a supreme spiritualisation, and this to the end that human beings might encounter God within a world that had become fully spiritualised through their own union with God. The world is created as a field where, through the world, humankind's free work can meet God's free work with a view to the ultimate and total encounter that will come about between them. For if humanity were the only free agent working within the world, it could not lead the world to a complete spiritualisation, that is, to its own full

encounter with God through the world. God makes use of humanity's free work within the world in order to help humanity, so that through humanity's free work both it and the world may be raised up to God and so that, in cooperation with humankind, God may lead the world toward that state wherein it serves as a means of perfect transparency between humanity and himself.[22]

Humans are not called to simply govern and organise creation in a worldly fashion. Instead, they are given the higher task of uplifting creation to complete spiritualisation. There is a remarkable convergence between this kind of theological language and the language used by some of the most prominent prophets of AI and the singularity. Futurists like Ray Kurzweil[23] or James Lovelock,[24] for example, believe that the cosmos longs for informatisation and that only future cyborgs or robots will be capable of saturating the universe with intelligence. Humanity's role, in their view, is that of a midwife to superior, synthetic forms of intelligence that will expand to corners of the universe inaccessible to biological life. Is this informatisation of matter the same as the spiritualisation invoked in Christian theology? I think not.[25]

Firstly, information does not equal spirit, despite both pointing to something immaterial. Nowadays, there exists this tendency to believe that anything that transcends the material domain must be informational. For example, the soul or the mind is sometimes regarded simply as informational pattern, which explains why some people in the transhumanist movement believe their minds could be uploaded to a computer. The Christian notion of spirit is much richer than the idea of information, pointing to a transcendent dimension of reality. Secondly, as evident in Stăniloae's account, Christian theology embeds

22 Dumitru Stăniloae, *The Experience of God: The World: Creation and Deification* (Holy Cross Orthodox Press, 2000), 59 (slightly altered).

23 Kurzweil, *The Singularity Is Near*, 21.

24 James Lovelock, *Novacene: The Coming Age of Hyperintelligence* (Penguin UK, 2019).

25 I argue this in more detail in Marius Dorobantu, "Why the Future Might Actually Need Us: A Theological Critique of the 'Humanity-As-Midwife-For-Artificial-Superintelligence' Proposal," *International Journal of Interactive Multimedia and Artificial Intelligence* 7:1 (2021): 44-51, https://doi.org/10.9781/ijimai.2021.07.005.

the spiritualisation of matter in the love relationship between God and humans. Spiritualising the world is not an end in itself, but rather a means to achieve complete transparency between creator and creation. Without God's love and purpose for creation, any spiritualisation/informatisation of matter is empty of content and significance. What would be the finality of such a process? A state of perfect and eternal cosmic equilibrium? In physics, such a scenario is known as the "big freeze," and it is synonymous with a heat death of the universe, where nothing more can happen due to a state of maximum entropy.[26] How could this be a cosmic state we should be rushing towards?

The theological account of the mystical role of humans in the world seems thus much more cogent than its secular counterparts. For theological anthropology, the implication is that a functional interpretation of the image of God needs to focus more on the spiritual dimension of our dominion and stewardship and not so much on its historical side, where AI may indeed outmatch us. The other dimension that needs to be stressed more concerning our vocation is the relational one. Our role in creation should not be divorced from our relationship with God. Being in the image of God does not entail just having been elected as God's representative at a certain point in or outside history. Instead, as shown by Stăniloae, it involves a continuous personal, authentic relationship of love between creature and creator, which brings us to the relational interpretation.

Vulnerable God, Vulnerable Humans, and the Image as Relationship

In a relational interpretation, the divine image is to be found in the loving relationships we develop with God and each other. Profound relationality is the mark of human life. We are born as a result of relationships, our personhood can only develop in relationships, and it is mostly in our relationships that we find meaning, purpose, and

26 A. V. Yurov, A. V. Astashenok, and P. F. González-Díaz, "Astronomical Bounds on a Future Big Freeze Singularity," *Gravitation and Cosmology* 14:3 (2008): 205–212, https://doi.org/10.1134/S0202289308030018.

fulfilment. If it is through relationships that we best mirror God, then developments in AI might legitimately question our distinctiveness. What if machines become one day capable of personal relationships? We already have conversations with chatbots, and the complexity of these conversations only increases as technology gets better. It is not unimaginable that in the future, we might talk to machines as we currently talk to humans.

This is precisely what Alan Turing proposed as a litmus test for whether a machine is truly intelligent. If someone conversing via text with the AI cannot tell whether they are talking to a human or a machine, then that machine should be considered intelligent.[27] This has become known as Turing's test and is still widely regarded as a valid benchmark for human-level AI. As of today, no program has passed the test, but as shown earlier, many people believe it to be just a matter of time before it happens. Would an AI capable of human-level conversations really engage in personal, authentic relationships? This is a tricky question, as shown by the confusion and heated debate that recently ensued when a Google engineer publicly expressed his concern that LaMDA, an AI he was working with, had become sentient.[28]

On the one hand, there are good reasons to believe that just displaying relational-like behaviour does not mean that an authentic relationship is actually being formed. Intuitively, a genuine self or consciousness is needed for the I-Thou type of relationship. Humans are such selves, while inanimate objects are not. Humans are *someone*, while machines are *something*. In Ted Peters' words, "nobody is at home" inside these machines.[29] On the other hand, we lack a convincing scientific theory to explain this difference between the presence and absence of consciousness, phenomenal experience, or subjectivity. In other words, we do not really know what makes us persons and

27 Alan M. Turing, "Computing Machinery and Intelligence," *Mind*, stb, 59:236 (1950): 433–60, https://doi.org/10.1093/mind/LIX.236.433.

28 Nitasha Tiku, "The Google Engineer Who Thinks the Company's AI Has Come to Life," *Washington Post*, November 6, 2022, https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

29 Ted Peters, "Will Superintelligence Lead to Spiritual Enhancement?" *Religions* 13:5 (2022): 5, https://doi.org/10.3390/rel13050399.

conscious agents. What is the secret ingredient that we possess and that robots lack? In the philosophy of mind, this is famously known as the "hard problem of consciousness,"[30] namely, how can consciousness or subjective experience arise from inert matter? Theologically, this problem can sometimes be dismissed with more ease if we believe in the existence of an immaterial soul. A supernatural soul could be a convenient explanation for the hard problem of consciousness. But unless one commits to a strong form of mind-body dualism that is at odds with most contemporary philosophy, speaking of a soul is plagued by the same kind of questions. Until we have a clearer understanding of what constitutes an authentic self, it is not wise to pontificate that machines will never become such selves.

People often point to the fact that AI is purely algorithmic and deterministic, thus incapable of consciousness, personhood, or freedom. But the same argument can be turned against humans because, from a scientific/mechanistic perspective, we are also algorithmic and, to some extent, deterministic beings, with the only difference that our algorithms are biological, genetic, or neurological, rather than digital or electronic. I do not necessarily subscribe to this view, but it is indeed tough to argue against it on purely scientific grounds. Insofar as the natural sciences are concerned, both humans and computers are machines, just different types. One needs to look at the issue from a completely different vantage point, for example of theology, to see something truly special about human beings. For the reasons listed above, it would be tough to decide whether an AI that acted as if it were conscious really was, or whether it was just simulating it. A robot claiming to be in love, suffer, or believe in God would pose challenging ethical, philosophical, and theological problems.

I think that contrary to what sci-fi likes to depict, the above scenario is improbable. AI is currently developing to *think* very differently from how humans think. When labelling images, playing Go, or responding to text messages, a human and a computer program might

30 David Chalmers, "Facing Up to the Problem of Consciousness," *Journal of Consciousness Studies* 2:3 (1995): 200–219.

sometimes produce the same result, but with very different tools and methodologies. Even when AI manages to attain human-level competency in various domains, it does so in a very non-humanlike fashion. When it makes mistakes, those are not the kind of mistakes that any human would make. Even if we somehow managed to endow our artificial creatures with a self and phenomenal experience, those would likely be radically different from our own due to our very different types of embodiment. Robots would have different perceptual senses, a different kind of access to their memories or internal states, and a very different relationship with time. Their needs would differ from ours, significantly impacting their interests and motivations. AI might indeed reach human-level competency someday, but it will probably be very non-humanlike.[31]

This is good news for the relational interpretation of *imago Dei* because it means that the kind of personal relationships that we have with each other will not necessarily be part of the robots' behavioural repertoire. Our relationality is very much connected with our vulnerability. We engage in relationships precisely because we are vulnerable and mortal, and need one another. There can be no genuine relationship without the two partners making themselves vulnerable to each other beyond any transactional logic. This is why deep relationships are always risky, because of the looming possibility of getting hurt. But without such voluntary vulnerability, how could anything deep and meaningful ever emerge? How could an artificial being, which is virtually invulnerable and immortal—having copied backups of itself on multiple computers—engage in humanlike relationships?

In Christian theology, this powerful idea that vulnerability is instrumental for authentic relationality is manifest in the doctrine of the incarnation. God does not shy away from vulnerability, but quite the contrary. Through Jesus Christ, we see God subjecting Godself to the ultimate vulnerability of suffering and death on the cross out of love

31 Marius Dorobantu, "Human-Level, but Non-Humanlike: Artificial Intelligence and a Multi-Level Relational Interpretation of the Imago Dei," *Philosophy, Theology and the Sciences (PTSc)* 8:1 (2021): 81–107, https://doi.org/10.1628/ptsc-2021-0006.

for creation. As humans, we image God when we are loving and vulnerable, not when we are mighty and unbreakable.

Besides vulnerability, another reason why human-level AI will likely not share in the kind of personal, humanlike relationships is its hyper-rationality. It is unlikely that a creature who makes all its decisions based on cold calculations of optimal outcomes will engage in such risky and irrational behaviour. We humans seek relationships because we have a sense of incompleteness and deep hunger for a kind of fulfilment that cannot be achieved solely within ourselves. Unlike the AI, we do not entirely understand our internal states and motivations, so we try to know ourselves better in relationships with others. That incompleteness drives us to seek the companionship of other humans, and it is arguably one of the main drivers of our religiosity, of why we seek God. This restlessness of our hearts, as Augustine called it,[32] or what Wolfhart Pannenberg refers to as exocentricity,[33] comes from deep within ourselves, from way below our rational minds. A purely rational being would not behave like this. Falling in love is certainly not a rational thing to do. However, it is such irrational things, from love to art to spirituality, that make human life enjoyable. Perhaps it is precisely because we are *not* as intelligent as AI that we can image God relationally.

The exciting developments in the field of AI arguably represent a blessing in disguise for theological anthropology, and this also constitutes an opportunity for a science-engaged theology. Far from endangering human distinctiveness, AI helps us appreciate some of the things that make us human and, therefore, different from machines. Following Aristotle, many Church Fathers believed that it is through our rationality and reason that we image God because that is what distinguishes us from the animals.[34] What reflection on AI shows is that,

32 "You have made us for yourself, and our heart is restless until it rests in you." Saint Augustine, *Confessions* (Oxford University Press, 2008), 3.

33 Wolfhart Pannenberg, *Anthropology in Theological Perspective* (Westminster Press, 1985), 51.

34 For example, Saint Gregory of Nyssa, *On the Making of Man* (CreateSpace Independent Publishing Platform, 2013).

although we might be more rational than nonhuman animals, we are certainly not the apex of rationality. Furthermore, because we are not entirely rational, we can engage in authentic relationships with other human persons and with God. In doing this, we mirror God, our creator, and become and flourish as authentic persons. Humans might look irrational and outdated when compared to the AI. Still, it is precisely because of our bodily and cognitive limitations that we can cultivate our divine likeness through loving, authentic, personal relationships. If reflecting on AI teaches theologians one thing, it is that our limitations are just as important as our abilities.[35] We may be vulnerable, but in being so we resemble a vulnerable God.

In my opinion, the truly ground-breaking conclusion from reflecting theologically on AI is that being like God does not necessarily mean being more intelligent. Christ's life and teaching show that what is most valuable about human nature are traits like empathy, forgiveness, and meekness, which are all eminently relational qualities. What enables such attributes is a kind of thinking rooted more in the irrational than in the rational parts of our minds. Perhaps this can shed new light on Saint Paul's choice to "boast all the more gladly of my weaknesses [...] for whenever I am weak, then I am strong."[36]

Conclusion

Although AI does not, in principle, challenge our theological understanding of human distinctiveness, our attitude towards this technology raises an important alarm about the future of human self-reflection. We are very much in awe of these machines and ready to consider them intelligent only until we understand how they work. In this sense, true AI has been an ever-receding horizon so far because our standards of what indeed constitutes intelligence are continuously shifting. John McCarthy, who coined the term *artificial intelligence*, says it best: "as

35 For a detailed argumentation, see Marius Dorobantu, "Cognitive Vulnerability, Artificial Intelligence, and the Image of God in Humans," *Journal of Disability & Religion* 25:1 (2021): 27–40, https://doi.org/10.1080/23312521.2020.1867025.

36 2 Corinthians 12: 9-10.

soon as it works, no one calls it AI anymore."[37] If we could travel back in time and show people fifty years ago the iPhone chatbot Siri, they would surely be astonished and consider it true AI. But for us, today, it is just another app. This is because we have looked behind the curtain, and we know more or less how it works: there is no magic involved! The more we understand how something works, the less inclined we are to ascribe it intelligence and value. This tendency is worrying because sometime in the future, it might be humans, instead of machines, that we disregard.

Our world is built around humanistic values, which stem from our fascination for the ultimate mystery of the human being. There are still so many things that we do not understand about ourselves, especially regarding our minds: what is the nature of thoughts, how are memories stored, how do we make decisions etc.? Human beings escape any complete theory or explanation, and this persisting mystery is probably one of the main reasons why we grant dignity and rights to human persons. Neuroscience and psychology are still in their infancy, but what if someday we *did* acquire a complete knowledge of the human person? What if we realised that we were, in fact, automata obeying algorithms that, although unspeakably complicated, are still ultimately deterministic? Should we then do away with human dignity and rights and treat humans as we currently treat other creatures and objects that we consider mindless? Obviously not. And this is precisely why theological anthropology should insist on an understanding of human distinctiveness and *imago Dei* rooted not in what humans are *on the inside*, as in the structural interpretation, but in our special relationship with God and the value of our relationships with each other. A move towards such a relational ontology would not only disentangle human dignity from human intellectual exceptionalism, but it would also arguably be more faithful to Christian Trinitarian theology.[38]

37 Moshe Y. Vardi, "Artificial Intelligence: Past and Future," *Communications of the ACM* 55:1 (2012): 5, https://doi.org/10.1145/2063176.2063177.

38 John D. Zizioulas, *Communion and Otherness: Further Studies in Personhood and the Church* (Bloomsbury Publishing, 2010).

Lastly, there is one area, in particular, where theological anthropology could bring a valuable contribution to the global discussion of our future with AI.[39] As Bostrom and many others have warned, there is a real danger in granting too much power to a technology whose control we could quickly lose. The worry is not that robots will consciously rebel against us like in the movies, but more that they might harm us involuntarily while trying to do exactly what we asked them to do. Concepts and values that would be obvious to a human being are not necessarily evident to a computer. That is why many brilliant computer scientists and philosophers are currently working on the so-called *AI alignment* problem. They try to ensure that even if machines eventually escape our direct control, their values will be sufficiently aligned with our own that they will not accidentally harm either us or anything else important to us. However, when it comes to which exact values to bake into these algorithms, things become complicated very quickly because there is no universal set of human values shared across cultures. It goes without saying that religious traditions should be part of this conversation because of the many people they represent and their centuries of experience reflecting on human values.

With all the noise generated by realising the potential threats of artificial super-intelligence, a more subtle danger goes completely unnoticed. Because most attention is devoted to preventing a catastrophic scenario, a consensus seems to uncritically emerge that an ASI that did not kill us would necessarily be good. We seem to be so caught into the otherwise crucial problem of aligning AI to our goals that we often do not even question whether we should even attempt to build ASI in the first place. The assumption is that it is good to bring about Bostrom's oracle/genie/sovereign because of the age of abundance, peace, and leisure that would follow. ASI would govern and feed us, take care of

39 For a broad discussion of issues in AI and Christian theology, see Marius Dorobantu, "AI and Christianity: Friends or Foes," in *Cambridge Companion to Religion and AI*, ed. Beth Singler and Fraser Watts (New York: Cambridge University Press, forthcoming); Marius Dorobantu, "Artificial Intelligence as a Testing Ground for Key Theological Questions," *Zygon*, https://doi.org/10.1111/zygo.12831.

our energy needs, and in general solve all the complex problems in our stead so that we could devote our lives to more pleasant endeavours. We would effectively be ASI's pets.[40] Who could possibly argue against such a future? How could the eradication of poverty and sickness not be a good thing? Although it is difficult to deny a certain appeal to this idea, many people would feel that something is just not right with this kind of *brave new world*. But this intuition cannot be articulated without an appeal to questions about what a good life is, the purpose of human existence, the value of vulnerability and suffering, and why freedom is ultimately more precious than comfort. To me, these are all theological questions and represent an exciting entry point for theology into the interdisciplinary and global dialogue on new technologies.

40 Samuel Gibbs, "Apple Co-Founder Steve Wozniak Says Humans Will Be Robots' Pets," *The Guardian*, June 25, 2015, sec. Technology, https://www.theguardian.com/technology/2015/jun/25/apple-co-founder-steve-wozniak-says-humans-will-be-robots-pets.