Artificial Intelligence and Religion: Recent Advances and Future Directions

with Andrea Vestrucci, "Introduction: Five Steps Towards a Religion–AI Dialogue"; Lluís Oviedo, "AI and Theology: Looking for a Positive—But Not Uncritical—Reception"; Christoph Benzmüller, "Symbolic AI and Gödel's Ontological Argument"; Sara Lumbreras, "Lessons from the Quest for Artificial Consciousness: The Emergence Criterion, Insight-Oriented AI, and Imago Dei"; Marius Dorobantu, "Artificial Intelligence as a Testing Ground for Key Theological Questions"; and Andrea Vestrucci, "Artificial Intelligence and God's Existence: Intersecting Theology and Computation."

ARTIFICIAL INTELLIGENCE AS A TESTING GROUND FOR KEY THEOLOGICAL QUESTIONS

by Marius Dorobantu

Abstract. Engagement with artificial intelligence (AI) can be highly beneficial for theology. This article maps the landscape of the various ways such engagement can occur. It begins by outlining the opportunities and limitations of computational theology before diving into speculative territory by imagining how robot theologians might think of divine revelation. The topic of AI and *imago Dei* is then reviewed, illustrating several ways AI can inform theological anthropology. The article concludes with a more speculative take on the possible implications of AI for divine infinity, incarnation, theodicy, and demonic intelligence.

Keywords: artificial intelligence; image of God (imago Dei); robotheology; strong AI; theological anthropology

---

Artificial intelligence (AI) has become a buzzword and can refer to many things. In this article, it is understood as the attempt to instantiate human cognitive abilities on artificial supports. Fundamentally, AI studies the nature of intelligence and whether it is possible to build machines that perfectly replicate or even outmatch human cognition. This capacity is of tremendous relevance for theological reflection.

One way of engaging AI theologically is by leveraging statistical algorithms' power to discern patterns in the vastness of our theological sources. Another possibility is to evaluate theologically our fascination with AI and

Marius Dorobantu is a Postdoctoral Research Associate and Lecturer at the Faculty of Religion and Theology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; e-mail: m.dorobantu@vu.nl.

 ISSN 0591-2385

see whether it is a sign of our greatness or a symptom of deep existential longing. However, another path is to ponder the future of AI and theologically assess the various possible scenarios. Could robots reach human-level intelligence? If so, what would this imply for human distinctiveness? Would AI also be in the image of God? Could robots become religious? If so, what kind of theology would appeal to them?

The Possibilities of Computational Theology

There is no explicit connection between contemporary AI and theology. The field of AI most often does not intentionally set out to explore anything that could be classified as theological. At best, AI is agnostic about theological issues, and at worst, it is atheistic, assuming a radical physicalism that excludes the existence of God, spirit, or even minds. This starkly contrasts with cybernetics, a precursor of AI, which was much more open to the acknowledgment of mystery in the world. Cyberneticians such as Norbert Wiener and Stafford Beer believed that humans must be more than mechanisms and that some things about the world and ourselves will always remain unknowable due to the cascading complexity of reality and the limitedness of our brains (Williams 1968, 44; Pickering 2004, 499–501). For them, the mystery of the divine did not come as something supplementary or superimposed but was seen to be in perfect continuation with the other unknowable aspects of the universe. Cybernetics was thus regarded as an exploration of this mystery. Such an explicit relation does not exist between religion and the successor of cybernetics, AI.

Employing AI programs to find hidden linguistic patterns in religious texts is perhaps the most straightforward and least speculative form of engagement between AI and theology. Computational methods have been used in biblical studies since at least the 1970s, but it was not until the advent of machine learning algorithms in the 2000s that the full potential of statistical AI was unlocked. Currently, computational methods are no longer an exotic approach in biblical studies but rather mainstream methodologies (Peursen 2017, 394). One example is how algorithms are helping biblical researchers distinguish between different authors in the same text (Dershowitz, Akiva, and Koppel 2015), something known as author clustering. The upside of leveraging the power of AI to study ancient texts is quite apparent: fresh insights, confirmation/disproval of hypotheses, and new connections. However, difficult *black-box*-type of problems arise when the program produces surprising results without being able to explain them. Should the researchers simply trust the AI to be right, which is unsatisfactory and arguably a slippery slope, or should they discard the results as mistakes and try to *repair* the algorithm until it produces the expected results, an approach that would, in turn, be circular and redundant? (Peursen Forthcoming, 11–12).

Things only get more complicated from here on. What if we could do much more with AI in the future than analyze texts? What if we could build algorithms that simulated catechumens well enough that religious teachers could test various pedagogical strategies on machines before applying them to human students? This is what computer scientist Edmund Furse proposed in 1986 (Furse 1986, 383). Given the impressive progress of natural language processing algorithms in the last two decades, such a possibility certainly looks plausible in the not-so-distant future, which could open exciting opportunities for missiology. However, Furse goes even further with his prediction. As our ability to computationally model human mental processes keeps improving, we might become able to build a computational model of a saint. Such models could then be tested in various real-life situations to see how a saint would behave.

Although tantalizing, this latter suggestion might be marked by a certain naiveté and overconfidence in the algorithms' ability to perfectly simulate human cognition, specific to the age when Furse was writing. This is obvious from his estimation that a computational model of Jesus would be "difficult to design," but not because of any insurmountable problem posed by the hypostatic union between the two natures, human and divine, as one would expect, but because of the insufficient training data available to the hypothetical algorithm (383). Arguably a more profound issue with building a computational model of a saint is the faulty circular logic behind the assumption. To model a saint, one would need to have either a complete and precise computational theory of holiness, which appears to be an absurdity because of the elusiveness of the concept, or a database of millions of examples of holy behavior carefully labeled for the AI to analyze and from which to generalize, supposedly, the heuristics of holy behavior. An obvious challenge of a logistical nature is that laborious labeling would still need to be done by humans. Who would even be qualified to do such a work? Another, more profound problem is the naiveté of the primary working assumption: that multiple humans could ever agree upon the criteria for holiness. This is highly improbable. If we could do that, we would not need a machine to help us understand holiness in the first place!

In addition to these "*technical*" obstacles, there might also be some theological problems with the dream of using AI to study perfect morality. As several authors point out (Weissenbacher 2018, 69; Samuelson 2020, 7), one relevant question is whether the fallen state of creation, in general, and of humans, in particular, would not hopelessly deter any attempt to instill holiness in machines. If sin is all-pervasive in the current state of creation, then AI could never completely bypass it: not because it would become itself evil, but more in the sense of inevitably reflecting its creators' moral shortcomings. To a certain extent, that is already noticeable in how

contemporary algorithms inherit their programmers' biases or those inherent in the training data.

Robots Doing Theology

Another enticing possibility is that advanced future AI could help us widen our understanding of divine revelation by providing us with a completely fresh perspective on some of the core tenets of religious faith. If robots ever reach human-level intelligence, then there might be a possibility that they will also become interested in religion. There are many caveats to this scenario. First, whether AI can, in principle, reach the human level is a hotly debated question without a clear answer thus far. Second, even if robots reach human-level intelligence, that does not necessarily mean they will also be humanlike *on the inside*. Robots might perfectly replicate humanlike behavior from an outsider's perspective, but their phenomenal experience of the world and themselves, if any, would be utterly different from our own experience of being a person in the world (Dorobantu 2021). Leaving these caveats aside, let us turn to a few hypothetical scenarios of religious robots and explore whether such thought experiments can shed some light on critical theological questions.

Philosopher Rajesh Sampath (Sampath 2018) tries to imagine how the Christian faith might be reinterpreted through the eyes of a hypothetical intelligent robot. Such a robot might understandably explore whether it, too, could be said to embody the image of God. The AI would therefore search for ways to interpret the New Testament and the core dogmas of the Christian faith as if they were written for and about robots. One way could be to think of Christ, the divine Logos, in terms of a software program and Christ's birth, death, and resurrection as akin to the program switching itself between ON and OFF. The pre-existence of the Logos would be understood simply as the eternal existence of the "Christ code" in God's mind. Could the Christ software program be *born* through an Immaculate Conception? Sure, if the latter is interpreted as the fact that the code was revealed at a particular moment in history when humans were culturally incapable of producing something like this. The point Sampath tries to make with these analogies is that an intelligent robot might be capable of coming up with its own original, although highly allegorical, theology, which would not contradict the Bible, nor would it breach the boundaries of Nicaean-Chalcedonian orthodoxy. Although humans and machines would agree on the *why* of the divine economy, they would see the *how* in markedly different terms. The critical question is why the robots' interpretation should be outright discarded in favor of the established human-centered account.

I am sympathetic to the importance of the question, although I have some reservations about how the argument is built. My main criticism

is that Sampath's working definitions of doctrine and divine revelation look slightly rigid. If revelation is seen as a finite collection of text (e.g., the Bible) and if orthodox doctrine is a finite set of logical propositions that can be exhaustively codified in symbols, then it might be true that these sets could, in principle, be manipulated in such a way to generate an interpretation that does not technically contradict the *rules of the game*. However, this is an impoverished account of both revelation and doctrine. In a more apophatic view, the content of faith is not exhausted in its linguistic expression, which can only approximate, at most, our religious intuitions. Taking the linguistic approximation and making it *the* ultimate reference point is theologically wrong. It is no wonder that it can lead to such forced interpretations as the one imagined by Sampath. His creative proposal might be technically correct in *letter* but arguably not in *spirit*.

However, the above criticism should not divert attention from the main point raised, namely, that it is, in principle, possible to imagine a radically different interpretation of divine economy than the one dominant thus far. History shows that Christian theology, for example, gradually extends to include the perspectives of formerly excluded categories—Gentiles, women, people of color—and intelligent robots can be regarded as the next *other* to lay legitimate hermeneutical claims. Even if this scenario might still be technologically far into the future, Sampath rightly pleads that it should serve as a reminder for striving toward a more inclusive pluralistic theology.

Although there is indeed a historical trend in the Christianity of expansion toward formerly excluded groups, the main drivers of this expansion are usually social and cultural rather than theological. Religious scholars Laura Ammon and Randall Reed compare the ecclesial debates over the inclusion of Native Americans in the sixteenth century and LGBT+ persons in our times. They aim to show that theological arguments play a surprisingly small role in reaching decisions about who is entitled to be a legitimate faith practitioner. The same will likely happen when robots become sufficiently advanced to raise the issue (Ammon and Reed 2019). If this is true, then the perceived adequacy of robot hermeneutics imagined by Sampath might depend less on orthodoxy and more on how integrated and accepted robots are in society at large.

Religious scholar James McBride tackles the same question from a different direction (McBride 2017). As we head toward a future where the presence of humanoid robots among us becomes ubiquitous, the eminently organic dimension of Christian theology—the one that refers specifically to flesh and the biological body—is likely to be increasingly questioned. Humanoid robots and their human partners/owners would find such an organic theology unappealing because of its insistence on God's incarnation, Christ's blood sacrifice, communion by eating and

drinking, baptism by total immersion in water, and the Pauline soteriology of the body. Instead, these AI "progressives" will expect theologies that give "meaning to all intelligent creatures, whether divinely or man-made" (671). McBride suggests that a solution to this potential future problem is a focus shift from the Pauline theology of the flesh (*sarx*) to Johannine *logos* theology. The latter would likely be more palatable to androids because they would identify better with the doctrine of a rational and intelligible universe due to the constitutional rationality of their source code.

Imago Dei after Darwin

Palatability to robots should probably not be the main criterion for preferring one theology over the other, or at least not yet. Nevertheless, McBride's argument illustrates how AI might start to play an increasingly important role in how we think about old theological questions. In this respect, the *imago Dei* debate is perhaps the one where the influence of AI is most noticeable. The case of *imago Dei* is particularly interesting because of its openness and high stakes. By openness, I mean that the jury is still out in regard to deciding what exactly is that renders humans in the image and likeness of God. By high stakes, I refer to the profound consequences of how we define *imago Dei* for how we think of the dignity of the human person and our role in the world.

Imago Dei is declared at the very beginning of the Hebrew Bible (Genesis 1:26 NRSV), but there is hardly any clue about how it should be interpreted. For many centuries, it was assumed that what makes humans like God and distinguishes them from animals is their intellect or some of its constitutive cognitive capacities, such as reason or the ability for language. Because of its emphasis on *something*—a capacity or the structure of the human psyche—that constitutes the image, this theology is known as the substantive or structural interpretation. As long as human superiority over the animals was self-evident, this interpretation went largely unchallenged. However, with Darwin and the advent of evolutionary theory, it became difficult to ground human distinctiveness on a purely ontological basis. We were suddenly not as different from the animals as we used to think. Moreover, it became evident that most of the intellectual abilities that rendered us distinctive had emerged naturally via evolution rather than having been bestowed upon us supernaturally by God. In the aftermath of this realization, theological anthropology has developed interpretations of the image that are arguably more nuanced and sophisticated.

Theological anthropology has undoubtedly benefitted from having to swallow the bitter pill of evolutionary theory, and the latter turned out to be a blessing in disguise in the long term. As theologian Aubrey Moore aptly put it, under the guise of a foe, Darwinism did the work of a friend (Ayala 2007, 159). This might be true more generally about theological

discourse: when it goes unchallenged for too long, it can cozy up to vague formulations and common places. In the long term, this usually leads to the weakening of its transformative power. Challenges—external, such as the Darwinian fundamental shift in scientific paradigm, or internal, such as the heresies of the first Christian centuries—can act as stressors in evolutionary parlance, shaking up the landscape of theological ideas and ultimately leading to better theological theory. Suppose this evolutionary view of theological epistemology is correct. In that case, future events, such as the emergence of human-level AI, could tremendously help our continual effort to refine our theological understanding of reality. McBride's hypothesis of a robot-determined future clash between Pauline and Johannine theologies is one example of how the story might unfold.

The impact of AI on *imago Dei* theology is another example. Precisely because post-Darwinian theology engaged so creatively with the challenges posed by evolutionary science, the result was a landscape where we now have too many appealing *imago Dei* interpretations to choose from: relational, functional, eschatological, Christological, and so on (Herzfeld 2002, 10–32). Although an excellent problem to have for theologians, this is nevertheless a problem. Each of these proposals is convincing in its way, capturing one or more of the aspects we intuitively associate with human distinctiveness and the divine image. Each is supported by biblical arguments and a cohort of brilliant theologians. On theological grounds alone, it is quasi-impossible to decisively turn the needle toward one interpretation or the other because we simply do not have enough constraint to prune out unlikely candidates: they all seem likely.

If theological arguments are undecisive, then perhaps comparing ourselves to our proximal relatives from the animal lineage can help us understand what it is that makes humans distinctive. However, this exercise also has strong limitations. Although evolutionary theory showed us that we were not so different in principle from nonhuman animals, in practice, we are still very different on multiple dimensions. This divide may be more related to quantitative/cumulative measures than to differences in quality or ontology. The fact remains, however, that anyway one looks at it, humans still end up looking pretty distinctive due to the richness of their cognitive arsenal, their individual and collective cultural achievements, and all the dimensions of life that are available to them to an extent unparalleled thus far in other animals: self-reflective, artistic, moral, spiritual, and so on. None of the major *imago Dei* interpretations is severely challenged by the comparison with the animals, so we are left with the same wide field of candidate proposals.

It is here, perhaps, that AI can make a difference (Dorobantu Forthcoming b). On the one hand, AI can be seen as challenging the final bastion of human distinctiveness by conquering precisely the cognitive capacities that we thought differentiated us from the animals.

We can ask, with mathematician John Puddefoot, "What, if anything, will remain of the 'uniquely human' when computer scientists (…) have done their worst?" (Puddefoot 1996, 83). On the other hand, AI can help us better understand our distinctiveness by indirectly shining a new type of light over us. One way in which reflection on AI can illuminate the mystery surrounding *imago Dei* is by deepening our understanding of the connection between the divine image and our creative effort to build intelligent machines. Another possibility is to analyze AI's achievements, failures, and opportunities and use AI as a reference for how we think back about human distinctiveness and what it means to be *imago Dei*.

What Can AI Teach Us about Ourselves?

Noreen Herzfeld and Anne Foerst—who both combine expertise in theology and computer science—believe that our attempt to build AI says something meaningful about the *imago Dei* in us. In her 2002 book on the topic, Herzfeld draws a remarkable analogy between the succession of paradigm shifts in AI and the evolution of *imago Dei* interpretations (Herzfeld 2002, 10–52). One conclusion she draws from this similarity is the depth to which implicit theological assumptions influence our technological endeavors. By trying to create AI in our own image, *imago hominis*, we unconsciously struggle to capture in machines what we think makes us distinctive and in the image of God (50). Ultimately, Herzfeld suggests, through AI, we attempt to create a distinct artificial *other* with whom we could relate. We are effectively trying to create a replacement for our lost relationship with God, the ultimate Other. This is why she thinks this attempt is doomed to fail because AI will never be capable of engaging in authentic relationships. Our obsession with AI, however, does reveal something important about ourselves, namely, the centrality of relationality and relationship for what it means to be human. This, according to Herzfeld, can be taken as evidence in favor of the relational dimension of *imago Dei* (51).

Foerst has a much more optimistic take on the possibilities of AI and what they reveal about human nature. She develops this argument in a pioneering 1998 article (Foerst 1998). Whereas Herzfeld saw in AI a deep human longing for relationality being misguidedly and hopelessly turned toward machines, Foerst has more sympathy for our effort to build AI. She regards it as a manifestation of the creative imperative inherent in the divine image, and she is ready to drop any claim that *imago Dei* qualitatively distinguishes humans from either animals or machines (108). Instead, if we ever manage to create human-level AI, such a momentous event could finally heal us from any pretension of ontological distinctiveness. In Foerst's view, human ontology consists of the very same mechanisms that animate nonhuman animals and, one day, perhaps, intelligent robots too.

Such a mechanistic view of human beings may very well be at home among many of the transhumanists and AI gurus in Silicon Valley, but it sounds rather unusual coming from a Christian theologian. Nevertheless, Foerst manages to put a positive spin on this reductionistic anthropology. If what renders us in the image of God is not to do with anything in our nature but solely with the divine mandate for stewardship, then both humans and other sufficiently intelligent creatures, including robots, could consciously choose to perform the *imago Dei*. Theologian Karen O'Donnell (2018) picks up this point and argues, drawing on Alistair McFadyen's theology, that advanced AI might lay legitimate claims at *imago Dei*, understood in such performative terms. For Foerst, thus, human nature is not something we are born with but a vocation, something we continuously create, never finished, and not exhaustively accounted for by either the Genesis narrative or the contemporary creation of AI. Instead, the two symbolic existential stories—*imago Dei* and AI—are complementary.

Although many theologians would hesitate to subscribe to such a mechanistic view of the human being, Foerst thinks of it as a liberating move. Realizing how much we share with nonhuman creatures invites, in her opinion, a more compassionate, responsible, and inclusive attitude toward nature, which brings us closer to the kind of stewardship mandated by the image of God we bear.

The performative view of *imago Dei* does have a certain appeal because of its potential to include nonhuman creatures. However, a question arises whether sufficiently intelligent robots could be labeled as *imago Dei* simply by virtue of their external performance. Theologian Jordan Wales thinks that AI built under the current paradigm could never aspire to authentic personhood, which requires consciousness and an interior life (Wales Forthcoming). The AI developed thus far has made significant progress toward intelligent behavior, but none we know of toward subjectivity and sentience. Current algorithms arguably lack any form of interiority. In theologian Ted Peters' words, "nobody is at home" (Peters 2022, 5). Without such interiority, thus without an authentic personal self, AI would remain ontologically *something*, and that is hardly enough for imaging a personal God.

It is not clear whether artifacts could ever acquire consciousness, understood here in philosopher Thomas Nagel's phenomenal sense: "an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism" (Nagel 1974, 436, emphases in original). We currently do not understand why humans are conscious in the first place, and, more generally, we do not have any convincing theory to explain how conscious experience can emerge out of inert matter, something philosopher David Chalmers famously dubbed "the hard problem of consciousness" (Chalmers 1995).

Until we develop such a theory, if ever, it is impossible to speculate on the theoretical possibility of artificial consciousness. However, even if that were possible, a scenario known as "strong AI" (Searle 2009), such an entity would still not very easily qualify as *imago Dei*, as I argue somewhere else (Dorobantu 2021), because it would likely be extremely nonhumanlike with respect to its feelings, thoughts, and perception of the world. Strong AI might behave in a humanlike manner to make itself understandable to us, but it would experience the world in a completely different way from biological organisms. With very different types of senses and needs, with complete access to its internal memory and states, and with the subjective passage of time slowed down a couple of thousand times, it is likely that an intelligent robot would think in categories that we cannot even imagine. Despite its consciousness and intelligence, I argue that it is unclear whether strong AI would also be a person because of its radical strangeness. Any discussion of AI becoming *imago Dei* is thus unwarranted, at least until we obtain a better understanding of the philosophical categories applicable to intelligent robots.

One thing Foerst did get entirely right about AI was the crucial link between intelligence and embodiment. Her study was realized while working with the team of roboticists at MIT who built the humanoid robot, Cog. The Cog project was the apotheosis of behavior-based robotics, an approach pioneered by Rodney Brooks, which was welcomed enthusiastically as a potential solution to the problems that plagued symbolic AI. The need for embodiment was one of the fundamental paradigm shifts proposed by Brooks & Co., in strong opposition to the disembodied kind of intelligence attempted in classical AI. Although some of the underlying assumptions in the Cog project were overly optimistic regarding how close Cog was to attaining human-level intelligence (Dorobantu Forthcoming a), the turn toward more embodiment proved to be directionally correct. In the three decades since, it has become increasingly clear that the insistence on disembodied intelligence, held by many in AI, was one of the things hampering progress in the field. Computer scientist Melanie Mitchell places it among the four common fallacies in our thinking about AI (Mitchell 2021).

According to Mitchell, the idea of disembodied intelligence can be traced back to mid-twentieth century psychology, but one could go even further to medieval mind/body dualism. From a philosophical and theological perspective, this dualistic anthropology starkly contrasts with the holistic, embodied, relational, and social biblical view of the human person. If anything, the history of AI and neuroscience thus far has proven that biblical holistic anthropology is a much more robust account than mind/body dualism (Barbour 1999). This is an emblematic example of how AI can serve as a testing ground for theological questions. Various competing approaches in AI are based on very different philosophical and,

one could say, even theological assumptions. Their successes and failures relative to each other can be used, to a certain extent, as some measure of the truth of those assumptions.

The underlying connection between implicit theologies and AI paradigms might look surprising, but it arguably goes very deep. Religious scholar Robert Geraci makes a compelling case that the kind of AI one is interested in building is strongly influenced by the theological assumptions imbuing one's culture (Geraci 2006). It is not accidental that computer science in the West (e.g., in the United States) has been historically more interested in disembodied AI, while in the East (e.g., in Japan), the focus is noticeably more on robotics. According to Geraci, this peculiar difference can be traced back to particularities in the religious traditions in which the two cultures are rooted. Eastern religions, for example, do not share the strong Western tabu for the ontological distinction between artificial and natural. Instead, these categories are blurred in East Asian cosmologies, where it is possible to see robots as participating "in a fundamental sanctity of the natural world" (229). The Western preference for disembodied AI over humanoid robots could be similarly explained through the prism of Christian eschatology. Although the latter never excludes the body, the emphasis is always on the salvation of the soul, which restarts its existence in a transfigured body (230). An unconscious connection should not be ruled out between this vision and some transhumanists' dream of uploading their minds into a computer simulation (235), where they could take up not one but multiple *transfigured* avatars of their choice.

AI and Some Wild Theological Speculations

This article ends with a short section that engages theology and AI more playfully. This approach is highly speculative and, thus, limited in its ambition. The purpose is to tease out some exciting possibilities for how paying attention to AI might result in unexpected insights into some of the most difficult theological questions.

The first such example is the notion of divine infinity. Monotheistic religions have always maintained that God must be infinite in all respects. However, as computer scientist Yorick Wilks argues, in a computational view of reality, that might not need to be the case (Wilks). Despite the vastness of the observable universe, spanning multiple orders of magnitude upward and downward from the human scale, it is still possible to characterize it in finite, albeit massive, numbers. The argument goes that the universe might be big but not so big to escape numerical characterization. If we can conceive of such large numbers, they are also computable on a colossal but finite computer. If God is creating and governing our world through some sort of computational process, Wilks boldly concludes, then God does not need to be infinite to do it. Just big enough would suffice.

From our perspective, it would make absolutely no difference because we would perceive God as unlimited for all purposes, but that does not preclude God from also having limits.

This idea is also present in the popular secular theology proposed by philosopher Nick Bostrom as the "simulation hypothesis" (Bostrom 2003). As we start contemplating the possibility of acquiring so much computing power that one day we will be able to simulate entire worlds, together with the sentient beings populating them, the thought arises that we too could be inhabiting just such a simulation created by more advanced beings. We could be one of their countless simulations. Or, even more wildly, our simulators could, in turn, be simulated by others and so on, until following this potentially colossally long chain finally leads to the only actual reality supporting all this dream in a dream. Bostrom is right to ask rhetorically what the likelihood is that our world is precisely that one original world, as opposed to just one of the many simulated ones. However, his whole argument hinges on the assumption that it *is* possible to simulate an entire universe, together with its fully-sentient beings. It is the latter, the part about the sentient inhabitants, that raises the most significant question mark because we do not know whether artificial sentience is theoretically possible. Are humans someone else's simulated artificial intelligences? The future of AI will hopefully shed some light on this question and indirectly contribute to our understanding of divine infinity/finiteness.

The recurrent idea in discussions about *imago Dei* and the simulation hypothesis is that by trying to create AI, we are in a somewhat analogous position to God's work at our own creation. Humanity's ultimate dream is to build strong AI, robots endowed with consciousness, volition, and freedom, just like us. However, in attempting to create an entity that is simultaneously pre-programmed *and* free, we might be able to glimpse God's dilemma when making us: how can you create an entity that is free when you are responsible for every ingredient, instruction, and process that goes into it?. The similarity between the two stories goes further. It is unclear how we could even measure whether our creation is conscious. All the possible tests we can conceive for AI, including the Turing test, evaluate competence and external behavior. It might be impossible to know whether intelligent robots are also sentient. Puddefoot speculates that this might be one of the reasons for divine incarnation: there was no way for God to know what it is like to be a human without becoming one (Puddefoot 1996, 123). Looking at the incarnation as God's ultimate Turing-like test for human consciousness, a verification that we are not, in fact, soulless zombies, is a very provocative idea that might get much theological pushback. However, it is a brilliant example of how the topic of AI can provide us with fresh perspectives on core theological issues.

In the same book, Puddefoot comes with another intriguing proposal. He suggests that reflecting on the possibilities of AI could bring us closer

to a solution to the notoriously difficult problem of theodicy (89). One typical response to why there is so much suffering in a world created by a good God, also known as the "only way" argument (Southgate 2008, 16), is that pain was the only possible way for God to bring about creatures endowed with high intelligence, such as ourselves. In other words, suffering at evolutionary scales is the price that needs to be paid for the emergence of intelligent creatures. However, if it becomes clear that computers can reach human-level intelligence and therefore do everything a human does, the above argument becomes problematic. If it is possible to bring to existence intelligent sentient beings by means of sheer programming, then why did God not choose this easier informatic path to bring about a world imbued with intelligence?

There are various possible answers, but Puddefoot's choice is to discern from this an alleged necessity for pain as a condition for flourishing, a sort of underlying law of our universe. Suffering must be valuable in itself. Otherwise, God would not have allowed it to play such a prominent role in the evolution of life in the universe. The conclusion can be stretched even further. Suppose suffering is so crucial for intelligence and development. In that case, the implication for AI is that to breach the gap toward humanlike intelligence, a robot "would need to grow, feel pain, experience and react to finitude, and generally enter the same state of mixed joy and sorrow as a human being. In particular, it would need to be finite, aware of its finitude, and condemned one day to die" (Puddefoot 1996, 92). This would be yet another confirmation that embodiment and phenomenal consciousness are *sine qua non* conditions for the emergence of humanlike intelligence and personhood.

In addition to the beautiful theologies arising from contemplating AI explored thus far, there is also a more sinister side to these technologies. There is an eerie resemblance between the way social media algorithms, powered by AI and big data, surveil and manipulate their users, and how demonic agency is described in traditional theology. As the story goes, demons follow humans around, observing them and trying to learn their weaknesses, to present them with just the right personalized temptation. Similarly, AI algorithms follow us around the Internet and *observe* every action we take online, trying to *understand* what excites or enrages us to be able to present us with just the right personalized content. As demons learn in time what works and what does not, so do the algorithms. Through A/B testing and profiling, they attempt to predict our behavior by continuously refining their model of who we are based on how we react to the information fed to us. Did we click? Did we buy? Did we scroll for another minute?

Ultimately, both demons and social media algorithms try to manipulate our behavior and persuade us to do things we would not necessarily do otherwise. Their purposes are, of course, different: while demons aim

to corrupt human souls, the algorithms are aimed at maximizing our engagement, time on the platform, or spending. Although the damaging effects on the human psyche might be the same in both cases—addiction, alienation, disinformation, guilt, moral corruption—in the latter case, we cannot speak of an evil intention *per se*, but just of side effects of the corporate greed behind the deployment of such harmful technologies. This similarity also does not mean obscuring the real evil behind such algorithms, which is ultimately humans exploiting other humans. AI algorithms are not (yet) agents endowed with volition, understanding, or purposes in the same ways humans are. Speaking of them *wanting*, *learning*, or *manipulating* is inevitably anthropomorphic, and the metaphorical nature of such language should always be kept in mind. Even though only symbolic, the similarity between some forms of AI and demonic intelligence is still striking.

The story might still contain a theological silver lining despite this bleak landscape. The fact that we are such easy prey to online algorithms means bad news for the mighty devil. If mindless and relatively simple programs can be so effective in manipulating us, then a personal hyperintelligent evil entity, as the devil is traditionally conceived to be, would have broken humanity to pieces long ago. Whereas social media algorithms can only collect information about us while we are online, the devil theoretically has 24/7 access to everything we do on and offline, to all the possible physiological data, and to a database composed of all the humans who have ever lived. Given all this computational arsenal supposedly available to demons, we can conclude that if such evil spirits exist, they must be pretty poor at their job and not particularly clever. Think only what the Facebook algorithm could do with such an informational treasure! A more pertinent conclusion would be that, if ever in doubt between a symbolic interpretation of the devil and a more literal, ontological one, theological conversation with AI seems to strongly point toward the former.

Even in its ugliest instantiation, AI can still bring good news to theologians.

Acknowledgments

This work 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.

I am indebted to Andrea Vestrucci, Philip Barnard, William Clocksin, Lluis Oviedo, Michael Reiss, Beth Singler, Fraser Watts, Yorick Wilks, Rowan Williams, and Harris Wiseman for their comments and suggestions.

References

Ammon, Laura, and Randall Reed. 2019. "Is Alexa My Neighbor?" *Journal of Posthuman Studies* 3 (2): 120–40.

Ayala, Francisco J. 2007. *Darwin's Gift to Science and Religion*. Washington, DC: Joseph Henry Press.

Barbour, Ian G. 1999. "Neuroscience, Artificial Intelligence, and Human Nature: Theological and Philosophical Reflections." *Zygon: Journal of Religion and Science* 34 (3): 361–98.

Bostrom, Nick. 2003. "Are We Living in a Computer Simulation?" *Philosophical Quarterly* 53 (211): 243–55.

Chalmers, David. 1995. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2 (3): 200–19.

Dershowitz, Idan, Navot Akiva, Moshe Koppel, and Nachum Dershowitz. 2015. "Computerized Source Criticism of Biblical Texts." *Journal of Biblical Literature* 134 (2): 253–71.

Dorobantu, Marius. 2021. "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): 81–107.

Dorobantu, Marius. Forthcoming a. "AI and Christianity: Friends or Foes?" In *The Cambridge Companion to Religion & Artificial Intelligence*, edited by Beth Singler and Fraser Watts. Cambridge: Cambridge University Press.

———. Forthcoming b. "Theological Anthropology Advanced by Artificial Intelligence." In *Progress in Theology*, edited by Gijsbert van den Brink, Rik Peels, and Bethany Sollereder. Oxford: Oxford University Press.

Foerst, Anne. 1998. "Cog, a Humanoid Robot, and the Question of the Image of God." *Zygon: Journal of Religion and Science* 33 (1): 91–111.

Furse, Edmund. 1986. "The Theology of Robots." *New Blackfriars* 67 (795): 377–86.

Geraci, Robert M. 2006. "Spiritual Robots: Religion and Our Scientific View of the Natural World." *Theology and Science* 4 (3): 229–46.

Herzfeld, Noreen L. 2002. *In Our Image: Artificial Intelligence and the Human Spirit*. Minneapolis: Fortress Press.

McBride, James. 2017. "Robotic Bodies and the Kairos of Humanoid Theologies." *Sophia*, 58: 663–76.

Mitchell, Melanie. 2021. "Why AI Is Harder Than We Think." Proceedings of the Genetic and Evolutionary Computation Conference 3.

Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *The Philosophical Review* 83 (4): 435–50.

O'Donnell, Karen. 2018. "Performing the Imago Dei: Human Enhancement, Artificial Intelligence and Optative Image-Bearing." *International Journal for the Study of the Christian Church* 18 (1): 4–15.

Peters, Ted. 2022. "Will Superintelligence Lead to Spiritual Enhancement?" *Religions* 13 (5): 399.

Pickering, Andrew. 2004. "The Science of the Unknowable: Stafford Beer's Cybernetic Informatics." *Kybernetes* 33 (3/4): 499–521.

Puddefoot, John C. 1996. *God and the Mind Machine: Computers, Artificial Intelligence and the Human Soul*. London: SPCK.

Sampath, Rajesh. 2018. "From Heidegger on Technology to an Inclusive Pluralistic Theology." In *Al and IA: Utopia or Extinction?*, edited by Ted Peters, 117–32. Adelaide: ATF.

Samuelson, Calum. 2020. "Artificial Intelligence: A Theological Approach." *The Way* 59 (3): 41–50.

Searle, John. 2009. "Chinese Room Argument." *Scholarpedia* 4 (8): 3100.

Southgate, Christopher. 2008. *The Groaning of Creation: God, Evolution, and the Problem of Evil*. Westminster John Knox Press.

van Peursen, Willem. 2017. *New Directions in the Computational Analysis of Biblical Poetry*. Brill.

———. Forthcoming. "Computational Linguistics." In *Linguistic Theory and the Biblical Text*, edited by William A. Ross and Elizabeth Robar. Cambridge Semitic Languages and Cultures. Open Book Publishers, University of Cambridge.

Wales, Jordan Joseph. Forthcoming. "Narcissus, the Serpent, and the Saint: Living Humanely in a World of Artificial Intelligence." In *All Creation Gives Praise: Essays at the Frontier*

of Science and Religion, edited by Jay Martin. Washington, DC: Catholic University of America Press.
Weissenbacher, Alan. 2018. "Artificial Intelligence and Intelligence Amplification: Salvation, Extinction, Faulty Assumptions, and Original Sin." In *AI and IA: Utopia or Extinction?*, edited by Ted Peters, 69–88. Adelaide: ATF.
Wilks, Yorick. Forthcoming. *God and Artificial intelligence* (*Preliminary Title*). Oxford: Oxford University Press.
Williams, Harold C. N. 1968. *Nothing to Fear*. Philadelphia: Pilgrim Press.