

Why the Future Might Actually Need Us: A Theological Critique of the ‘Humanity-As-Midwife-For-Artificial-Superintelligence’ Proposal

Marius Dorobantu*

Vrije Universiteit Amsterdam (Netherlands)

Received 3 February 2021 | Accepted 19 June 2021 | Published 30 July 2021



ABSTRACT

If machines could one day acquire superhuman intelligence, what role would still be left for humans to play in the world? The ‘midwife proposal,’ coming from futurists like Ray Kurzweil or James Lovelock, sees the invention of AI as a fulfillment of humanity’s cosmic destiny. The universe ‘strives’ to be saturated with intelligence, and our cyborg descendants are much better equipped to advance this goal. By creating AI, humans play their humble, but instrumental, part in the grand scheme. The midwife proposal looks remarkably similar to modern Christian anthropology and cosmology, which regard humankind as “evolution becoming conscious of itself” (Pierre Teilhard de Chardin), and matter as having a predisposition to evolve toward spirit (Karl Rahner, Dumitru Stăniloae). This paper demonstrates that the similarity is only superficial. Compared to the midwife hypothesis, Christian theological accounts define the cosmic role of humanity quite differently, and they provide a more satisfactory teleology. In addition, the scientific and philosophical assumptions behind the midwife hypothesis – that the cosmos is fundamentally informational, that it intrinsically promotes higher intelligence, or that we are heading toward a technological singularity - are rather questionable, with potentially significant theological and ethical consequences.

KEYWORDS

Artificial Superintelligence, James Lovelock, Novacene, Ray Kurzweil, Singularity, Theological Anthropology.

DOI: 10.9781/ijimai.2021.07.005

I. THE FUTURE STILL NEEDS US, BUT ONLY FOR A WHILE

AT the turn of the century Bill Joy, then Chief Scientist at Sun Microsystems, published in Wired his famous article, “Why the Future Doesn’t Need Us” [1]. It represented a moment of chilling public realization that some of the most dystopian future scenarios were no longer mere sci-fi fantasies. In Joy’s view, the convergence of robotics, genetic engineering, and nanotechnology could constitute an existential threat for humanity, rendering us an “endangered species.”

Today such an article would hardly make the headlines. But what is perhaps most remarkable about Joy’s piece is that it came at the end of a so-called AI winter, a period of declining hype, interest, and funding for AI, triggered by the field’s failure to deliver on its naïve and grandiose initial promises. AI research had emphatically taken off in the 1950’s, with resounding successes in mathematics and game-playing, things notoriously difficult for humans. Computer programs could solve problems, prove theorems, and play some strategy games at human-level performance. Since such highly intellectual tasks proved relatively easy to replicate, it was widely believed that more mundane abilities, like perception or locomotion, would not pose too many problems. This is well illustrated by how MIT scientists in the 1960’s thought that they could collectively solve computer vision as a summer project [2]. Even more ambitiously, it was predicted that within just one

generation we would witness the first machine endowed with human-level intelligence [3]. It is needless to say that that wasn’t the case. The ‘mundane’ capabilities of human intelligence proved much more difficult to replicate than the ‘intellectual’ ones, something known as Moravec’s paradox [4]. The failure to deliver on those big promises deeply affected the public perception of AI, and the possibility of truly intelligent machines was again relegated to the realm of fantasy and sci-fi. This mood was still in place in the late 1990’s and early 2000’s, around the time when Joy’s article saw the light.

However, unbeknown to the public, new technologies were developing, and a new AI-revolution was brewing. The first wave of AI had operated from the assumption that the human mind processes information sequentially, like a computer algorithm, by manipulating a finite set of symbols by means of logical operations. Teach a computer program enough of these symbols, rigorously define the rules for how they should be combined, and a powerful enough computer should start to reason like a human. To the contrary, the approach behind second-wave AI, generally known as machine learning, was rather different. Instead of assuming any theory of human cognition, it tried to roughly imitate the organ of human cognition, that is, the brain, and see whether this could lead to intelligent behavior. The most successful branch of this approach, deep learning, stacks many layers of artificial neural networks together and trains them to recognize patterns, presumably in a somewhat similar fashion to how a human brain learns to recognize patterns in the world.

Deep learning lies behind the biggest AI successes of the new millennium. Our banks, our stock markets, our airports, our

* Corresponding author.

E-mail address: m.dorobantu@vu.nl

smartphones, and our social media feeds are largely powered by second-wave AI algorithms. AI programs today are capable of learning from scratch how to play strategy games. They can compose symphonies and coherent texts, and they can also make decisions in spite of incomplete information. All these arguably resemble learning, creativity, or intuition, capacities once considered to be uniquely human. They are also critical in another respect: they open the possibility for AI to re-design itself and augment its intelligence. Concerns have been formulated that if AI were to reach human-level competency at programming AI, this could trigger a positive feedback loop. Such an AI could build a more capable version of itself, that would, in turn, be more competent at building an even better version and so on. This scenario is known as an “intelligence explosion,” a term introduced by I.J. Good [5]. AI would thus quickly reach superintelligence, a level way above all humans collectively.

Nick Bostrom masterfully demonstrates that such an artificial superintelligence (ASI) would be impossible to contain, unless we could insure from its inception that it is friendly towards us [6]. In AI, this is known as the famous ‘alignment problem,’ and is notoriously difficult: how to make sure that an ASI will have goals that are aligned with our own? The reality is that we have almost no way of anticipating how such an ASI might see us, or what its goals might be. It is sometimes said that it would be as intelligent compared to us as we are compared to ants. If this turns out to be true, then ASI would inhabit a very different perceptual and conceptual world, and it would perhaps have very different kinds of thoughts, that we couldn’t ever comprehend.

Whether such an ASI would be a ‘true intelligence’ or a mere automaton is of little relevance for the outcome of this scenario. John Searle famously distinguishes between ‘strong AI’ and ‘weak AI’: while the former would be capable of thinking and would have what we call consciousness, the latter would only be a perfect simulation of intelligence, with no subjectivity, thoughts or phenomenal experience [7]. Because we do not have an answer yet to the question of how physical matter can bring about conscious experience – something known as the “hard problem of consciousness” [8] –, we also do not know if ASI would be strong or weak AI.

Whether or not ASI would be a *someone* instead of *something* is a fascinating philosophical and theological question, but with little implication for how such a future might play out for us in practice. In fact, too much talk of artificial consciousness might be a red herring, distracting us from the real possibility that AI might soon outsmart us. The question whether AI can be conscious is different from the question whether it can be more intelligent – understood as competent – than us, and too much attention to the former might obscure the astonishing AI progress in the latter. It is a chilling realization that ASI might be possible without having consciousness and a mind, but current AI algorithms already achieve super-human levels in many tasks without possessing such things. For practical purposes, it doesn’t really matter whether there is a ‘real mind’ and intentionality behind an AI program, as long as it can outperform humans in increasingly more relevant cognitive tasks. Before too long, we might wake up to a reality where machines have reached human-level intelligence, or even ASI.

Talking now of ASI might seem exaggerated, but many experts with first-hand experience in the field believe that it is only a matter of decades until we AI gets there. An often-cited 2016 survey of 550 AI experts reveals that most of them expect human-level AI between 2040 and 2050, and ASI by 2080, the latest [9].

Machines are currently dependent on us to program and power them on, but this might cease to be the case with future, more autonomous, robots. Thinking of such intriguing scenarios helps us realize that a future without humans is perfectly possible, as Bill Joy eerily predicted. Some have reacted to this realization by ringing the alarm bells about

the stakes of AI research. In 2015, Stephen Hawking, Elon Musk, and dozens of AI experts signed an open letter, calling for more research on the societal impact of AI and against mindlessly building something that cannot be *a priori* well-understood, let alone controlled [10].

Others, however, seem to be more reconciled with a future that does not necessarily include humans. Our cyborg descendants are going to replace us and that is absolutely fine. In the natural world, it is normal for better adapted life forms to thrive and replace their progenitors, so eventually this was going to happen to our species anyway. The special thing about our demise, though, is that it will mark the transition from biological to artificial life. This is what is referred to in the paper as the ‘midwife proposal.’

There are two ways in which this transition can play out. In Ray Kurzweil’s account, humans merge with machines, giving birth to hybrid cyborg species [11]. In James Lovelock’s prediction, humans gradually fade away like an endangered species, being gradually replaced by intelligent forms of artificial life [12]. In either case, our hyper-intelligent descendants will go on expanding to other planets beyond the solar system, and then beyond the galaxy into the rest of the universe, saturating it with intelligence. This is something that we are not equipped to do, due the limitations of our biology. Humanity would thus be casted in the midwife role in this cosmic evolutionary drama.

II. KURZWEIL’S SINGULARITY

Inventor and futurist Ray Kurzweil predicts that AI will reach human-level by the year 2029, which will be demonstrated by the first program to pass the Turing Test [11, p. 200]. But there will be no reason for AI to stop at human level. Beyond that point, its capability would continue to grow exponentially, leading to a so-called ‘singularity’ in 2045. This is when humans would merge with machines into a new type of hybrid being, many orders of magnitude more intelligent than *Homo sapiens* [11, p. 136].

The notion of technological singularity is rooted in mathematics and physics, denoting a point of no return in history. Just as it is impossible to communicate information from beyond the event horizon of a physical singularity, also known as a black hole, so too it is impossible to predict what history will look like after the technological singularity. It was mathematician John von Neumann who first associated the concept of singularity with technological progress [13], while Vernor Vinge popularized it in the 1990s [14].

The main principle behind Kurzweil’s bold prediction is what he calls “The Law of Accelerating Returns.” He regards human history as a history of technological evolution, anticipating that technological progress will continue forward at an accelerated rate. This acceleration is to a certain extent already accounted for by Moore’s Law. In the 1960s, Intel co-founder Gordon Moore correctly predicted that the density of transistors in integrated circuits would continue to double at regular intervals [15], thus making computing technology exponentially cheaper and more powerful. But Kurzweil’s Law of Accelerating Returns goes beyond this, conjecturing that the exponential growth tendency applies to a wider variety of evolutionary systems [16]. In other words, Moore’s Law would only represent a particular case of the more general Law of Accelerating Returns, according to which technological progress in the world occurs at an accelerated rate.

The 2029 singularity would be followed by a complete merging between biological (human) and artificial intelligence. The resulting super-intelligent being would combine the best of each realms: “The Singularity will represent the culmination of the merger of our biological thinking and existence with our technology, resulting in a world that is still human but that transcends our biological roots.

There will be no distinction, post-Singularity, between human and machine or between physical and virtual reality” [11, p. 9].

Post-singularity cyborgs will supposedly combine the unique traits of human intelligence – the plasticity and massive parallelism of our brain, our mind’s ability to hold contradictory thoughts etc. – with the advantages of AI – the huge speed of electronic circuits, increased memory storage, the ability to instantly copy skills/programs from one machine to the other etc. These godly successors of ours will live lives that are incomprehensible to us and will have powers that we cannot even think of. They will proceed to fulfill the universe’s “ultimate destiny,” that is, to be infused with intelligence: “In the aftermath of the Singularity, intelligence, derived from its biological origins in human brains and its technological origins in human ingenuity, will begin to saturate the matter and energy in its midst. It will achieve this by reorganizing matter and energy to provide an optimal level of computation [...] to spread out from its origin on Earth. [...] Whether our civilization infuses the rest of the universe with its creativity and intelligence quickly or slowly depends on [the speed of light’s] immutability. In any event the “dumb” matter and mechanisms of the universe will be transformed into exquisitely sublime forms of intelligence [...]. This is the ultimate destiny of the Singularity and of the universe” [11, p. 21].

III. LOVELOCK’S NOVACENE

Futurist James Lovelock is best known for his Gaia hypothesis, which posits that the Earth is a self-regulating system, like a giant organism, and that the emergence of life is part of our planet’s evolutionary ‘strategy’ to keep cool against the increasing energy output of the Sun [17]. In his book, *Novacene: The Coming Age of Hyperintelligence*, Lovelock argues that we are quickly approaching the end of the Anthropocene and the beginning of a new geological age, the Novacene. The defining characteristic of this new age is the emergence of electronic life capable of directly transforming energy into information.

The main assumption behind Lovelock’s argument is that the cosmos is informational at its most fundamental level. This would explain what he sees as the consistent positive selection of intelligence throughout cosmic evolutionary history, leading to the emergence of humans, the first ‘understanders’ of the cosmos. The ‘informational assumption’ would also neatly explain the anthropic principle, namely, the apparent fine tuning of physical laws and constants for the emergence of intelligent life. If this is true, then humans are the first consciousness of the cosmos. Through us, the universe awakens and becomes self-aware. Through our hyper-intelligent cyborg descendants, the universe will accomplish its last evolutionary stage, that of transforming all its matter and energy into information [12, p. 123].

The Gaian super-organism is therefore a nursery for the cosmos’ self-awareness. Lovelock identifies three key moments in the history of Gaia, each corresponding to the beginning of one distinct geological age. The first was the evolution of organisms capable of photosynthesis, enabling Gaia to capture sunlight and store its energy. By releasing oxygen, these organisms set the stage for the emergence of more complex life, culminating with humans. The second key moment was the invention of steam-engines. Through technology, humans triggered the second stage, the Anthropocene, marked by Gaia’s capability to transform the stored solar energy from fossil fuels into useful work. The third stage, the Novacene, begins when humans invent machines that are capable of learning and re-designing themselves, with a widespread ability to transform sunlight into information. Later, these machines will pursue the conversion of all the physical matter into information.

In Lovelock’s story, these electronic life forms come in perfect continuation with biological life, emerging through the same processes of Darwinian selection. Instead of the natural selection that characterizes the evolution of biological organisms, cyborgs will undergo a purposeful selection, marked by a quick correction of harmful mutations.

Zooming out one further level, the Novacene can be regarded as a necessary evolutionary ‘strategy’ by Gaia. In a few hundred million years the Sun is poised to become a Red Giant type of star, dramatically increasing its radiation output. Biological life won’t be able to keep the planet cool in such conditions anymore, hence the need for super-intelligent electronic life forms, with technologies powerful enough to tackle this challenge.

How will the Novacene unfold? Similarly to Kurzweil, Lovelock speculates that the evolution of cyborgs will be accelerated. Differently from Kurzweil, he does not think that humans will be able to keep up. Speed is one main quantitative differentiator between human and artificial intelligence. Electrical current can travel much faster through electronic circuits than through a brain’s wetware, potentially 1 million times faster. Lovelock settles for the more conservative prediction that cyborgs will think around 10.000 times faster than we think [12, p. 81]. Even so, this would be the same ratio as that between humans and plants.

Another big difference, this time a qualitative one, would come from the very different nature of cyborg intelligence. AI will allegedly be more intuitive than human intelligence, because it wouldn’t be built around speech. As the story goes, humans developed speech as a necessary evil. While it has been of tremendous evolutionary value for our species, it has also narrowed our thinking to the current linear, step-by-step logic, that we are all familiar with. Cyborgs won’t have a speech-driven intelligence, and they will likely communicate telepathically with each other, retaining speech only to be able to communicate with us [12, pp. 96-103]. This will supposedly free up their intelligence from the chains of discursive thinking to realms and possibilities that for us are difficult to even imagine.

While it seems right to suppose that cyborgs will develop a very non-humanlike type of intelligence [18], the choice to label human intelligence as discursive and AI as intuitive looks rather odd. If anything, it should be the other way around. Modern psychology posits that there are more (usually two) modes of human cognition, of which only one is sequential and discursive: Daniel Kahneman’s system 1 and 2 [19], Philip Barnard and John Teasdale’s propositional and implicational [20], or Jonathan Haidt’s elephant and elephant rider [21], to name only a few. Iain McGilchrist masterfully demonstrates that the intuitive, right hemisphere type of intelligence is much more involved in human cognition than the rational, left hemisphere type [22]. If the history of AI so far is of any relevance, machines are actually more likely to master the discursive type of intelligence and to struggle with the intuitive one. Hubert Dreyfus’ famously argued that computers might have a chance of replicating our conscious, “knowing-what” mode of cognition, but it would incomparably more difficult for them to master the unconscious, “knowing-how” mode [23].

As for how these cyborgs will behave towards us, their ‘parents,’ Lovelock manifests an unbridled optimism, foreseeing no likely power struggle between biological and electronic life. Cyborgs will be so far ahead of us that we could not possibly be a match for them, just as other animals are currently no match for us. However, they will probably keep us around, in order to help keep Gaia cool. Future AI will be intelligent enough to realize that the biggest threat for its existence comes from the increasing heat from the Sun. The best strategy against that, at least for the time being, would be to preserve

biological life and allow it to continue cooling the planet, perhaps with some help from cyborg-invented technologies.

But is it realistic to believe that the overheating from the Sun, not due for a few hundred million years, would be so high on the immediate agenda of ASI? Would ASI not be able to invent better technologies for that job? Or, better, could the cyborgs not decide to keep all biological life *except* for humans? After all, no other species has ‘worked’ harder than we have against the temperature homeostasis throughout the Anthropocene. Lovelock is confident that our descendants will not bother to exterminate us, but rather treat us like pets, an idea going back to Apple’s co-founder, Steve Wozniak. In this scenario humans will therefore enjoy a “peaceful retirement” [12, p. 119].

How does one so easily come to terms with such a chilling possibility? The answer is that it largely depends on the value ascribed to human life and to a human presence in the world. In Lovelock’s system, the cosmos we live in is intelligence-oriented, but not necessarily anthropic-oriented. It manifests a selective preference for intelligent organisms and a predisposition toward being converted from matter to information. Up until now, humans have been the best at promoting that. But cyborgs will be far better equipped for this task than humans, so we should simply accept that “we have played our part” and rejoice “as wisdom and understanding spread outwards from the Earth to embrace the cosmos” [12, p. 130].

IV. INTELLIGENCE AS A COSMIC GOAL

Kurzweil and Lovelock’s visions differ slightly in the kind of cyborg descendants they predict. While Lovelock forecasts fully cybernetic organisms, Kurzweil envisages a hybrid between humans and AI along transhumanist lines, but even in such a case, these hybrids would arguably be anything but human. In both scenarios, humanity plays the critical, but historically limited, role of a midwife for the more advanced forms of cyber life.

At a more profound level, something that both these extreme versions of cheerful non-anthropocentrism have in common is the assumption that our universe has a built-in purpose to promote intelligence, ultimately understood as total informatization of matter and energy.¹ Both versions of the midwife proposal presuppose that the universe has a destiny, and that humans play a pivotal role in the cosmic drama of fulfilling that destiny. This assumption takes the midwife proposal from the realm of scientific ideas, where both authors claim to be dwelling, and transports it very close to the realm of the religious.

The idea that humans have a central role to play in the world is not very different from what most major religions are claiming. In Christian anthropology, for example, this is known as the functional interpretation of human distinctiveness and the image of God (*imago Dei*) [25]. In the functional view of the *imago Dei*, humans are being regarded as God’s representatives in the created universe, appointed to exercise stewardship and dominion over the world [26]. In the posthuman scenarios of Kurzweil and Lovelock, humans fulfill the role of midwives for the new hybrid intelligence that will “saturate the matter and energy in its midst” [11, p. 21]. This vocational language is strikingly reminiscent of God’s commandment to the first humans to “fill the earth and subdue it” [27].

¹ It is worth pointing out that from a strictly scientific perspective this claim is highly dubious, if not a category error altogether. Matter/energy and information are not interchangeable quantities, because they exist at different levels of conceptual ‘zooming-in.’ Matter and energy are to a certain extent equivalent, as shown by Einstein’s famous $E=mc^2$ equation, but information is a totally different level of characterization of reality. Matter cannot be ‘informatized,’ as Lovelock suggests. Moreover, if the information conservation principle is correct [24], then no new information can be created in the universe.

The jump from AI and technological singularity to Christian cosmology and the image of God may seem abrupt, but the two realms have much more in common than what the casual reader might initially think. Firstly, the two discourses operate with surprisingly similar notions and structures, such as prophetism [28], a dualistic view of the world, or transcendent promises for the future. Transhumanist views of the future and ‘AI apocalypticism’ draw substantively on religious thought, especially on Jewish and Christian apocalypticism, to the point where the former can reliably be traced back as a “legitimate heir[s]” to the latter [29]. They also share the belief in a clear periodization of future cosmic history, and in profound changes that are imminent for both humanity and the world [30].

Secondly, they are equally non-scientific. This is not intended as a derogatory judgement, but as a precise delimitation of the space in which the two discourses are meaningful. While modern theology openly admits its limited competence when it comes to scientific issues, secular ideologies, such as the midwife proposal, often present themselves as flowing naturally from scientific theory and observations of the world. But just as Christian theology departs from some clearly defined assumptions about the existence of God and the meaning of life, so too such ideologies make strong assumptions about the purpose of the universe that are eminently non-scientific in nature.

The relegation of humans to a position of evolutionary ancestors or midwives of the true “heirs of the world” [31], namely super-intelligent cyborgs, is a scenario fueled not so much by irrefutable scientific observations, as by Kurzweil and Lovelock’s personal beliefs regarding the teleology of the cosmos and the inevitability of the hybridization between human and machine.

From the conjecture that the universe is primed to favor intelligent forms of matter and energy, the ideology behind the midwife proposal goes on to affirm that the intrinsic purpose of the universe is to be saturated with intelligence. Biological life, of which humans are the apex, could never accomplish the cosmic goal. AI will therefore take over at some point and continue unhindered to use more and more resources to augment its intelligence until finally the entire cosmos becomes saturated with intelligence. As further shown, far from being neutral, these views are scientifically dubious, theologically problematic and morally dangerous.

V. THE MIDWIFE PROPOSAL IS SCIENTIFICALLY DUBIOUS

Both Kurzweil and Lovelock present their ideas as valid scientific paradigms that could account for our observations of reality. Lovelock’s Gaia hypothesis is allegedly nothing but Darwinian evolution extended to the cosmic scale, while Kurzweil’s Law of Accelerating Returns is a clever generalization of Moore’s law. However, their proposals are not as purely scientific as they are presented to be. Instead, as further shown, they are a complicated mixture between the scientific and the non-scientific, between good and crazy ideas, as pointed out by Douglas Hofstadter’s scathing criticism of Kurzweil: “[I]t’s a very bizarre mixture of ideas that are solid and good with ideas that are crazy. It’s as if you took a lot of very good food and some dog excrement and blended it all up so that you can’t possibly figure out what’s good or bad. It’s an intimate mixture of rubbish and good ideas, and it’s very hard to disentangle the two” [32].

The observation that our universe seems to favor intelligent life-forms may hold some validity, but it looks to be heavily biased toward the particular evolution of life on *our* planet. On a local scale, it seems indeed true that more intelligent life forms are often evolutionarily fitter. From this perspective, developing AI that potentially surpasses our intelligence may indeed be our worst evolutionary mistake, as pointed out by Stephen Hawking [33].

On a cosmic scale, however, the conjecture that the universe selects for intelligence hardly finds any support. If intelligence is indeed a great attractor to which the universe is irresistibly drawn, how come our observations so far have not revealed a ubiquity of intelligent forms of existence? The contrast between this expectation to find myriads of technological civilizations in the observable universe and the apparent silence in our observations is known as the Fermi paradox. In the summer of 1950, Italian physicist Enrico Fermi supposedly asked: "Where is everybody?" [34]. The universe is already approximately 13.8 billion years old [35]. Given its presumed predisposition to favor intelligence, one would expect that at least a few other life forms more advanced than our own have already gone through the technological singularity and/or have given birth to space-expanding AI. The process of saturating the universe with intelligence should thus be noticeably underway. But from what has been so far observed, it is not. This begs the question: why?

Both Kurzweil and Lovelock choose the same answer, which is intuitive, but astronomically improbable: we are the first ones [11, p. 357] [12, pp. 3-5]. We do not observe extra-terrestrial intelligences because there are none yet. But if the universe is capable of producing countless intelligent species throughout its existence, and even more if it is primed to do so, how likely is it that we are the first of this colossally big series?

The Fermi paradox is one of the most complex scientific and philosophical problems of our times, and its scope is simply too wide to be discussed in detail in this paper, so let us only point out that there exist a variety of proposed solutions. One of them, for example, formulated by Anders Sandberg, Eric Drexler, and Toby Oord, conjectures that we might be indeed the only technological civilization in the observable universe, but not because we happen to just be the first one, but rather because the emergence of intelligent life is incredibly improbable [36]. This looks like a more promising explanation, but it pushes back against the idea that the universe is 'striving' towards intelligence and informatization, which is a crucial assumption of the midwife proposal. How could only one occurrence of intelligent life across such a vastness of space and time be typical of anything?

Kurzweil actually acknowledges that his solution of humanity being the first intelligence of the cosmos has a dramatically low mathematical probability. But instead of embracing the 'improbability of intelligence' explanation, he chooses to invoke a modified version of the anthropic principle: some civilization *has* to be the first, and if our observations suggest that we are the first, then it must be true, in spite of the astronomical unlikelihood. This allows him to continue to postulate the built-in longing for intelligence of the cosmos, and the critical role of humanity in driving the universe toward "complexity and order:" "[W]e are in the lead. That's right, our humble civilization with its pickup trucks, fast food, and persistent conflicts (and computation!) is in the lead in terms of the creation of complexity and order in the universe" [11, p. 357].

Besides the casual romanticism in Kurzweil's above quote and the apparent bias to equate human civilization with the United States of America, it becomes clear that the midwife proposal also has many philosophical and theological implications.

VI. THE MIDWIFE PROPOSAL IS THEOLOGICALLY PROBLEMATIC

Even if the cosmos systematically promotes intelligent life-forms, an assumption shown above to be questionable, it seems a step too far to conclude that the universe has the goal of becoming more complex, ordered, or more saturated with intelligence. Firstly, if anything, the second law of thermodynamics paints the opposite picture of a

universe inevitably evolving from order to chaos on a global scale. Secondly, speaking of goals and meaning is usually not the province of science, but of philosophy and theology.

On a subtler level, in spite of its apparent radical non-anthropocentrism, the midwife proposal actually supports its own version of human exceptionalism: it is, after all, humans who have the seminal role of making the transition between biological and cybernetical stages of evolution. As already noted, this appears to bear some similarities with functional accounts of human distinctiveness in theological anthropology.

The idea that humanity represents an inflection point in cosmic history is not something new. In fact, it has been emphasized so much in theological anthropology, to the point of inviting accusations of anthropocentrism [37], [38]. The very notion that the universe 'awakes' for the first time in human beings is not strange to Christian evolutionary thought. Paleontologist and Jesuit priest Pierre Teilhard de Chardin refers to humankind as "evolution becoming conscious of itself" [39]. Roman-Catholic theologian Karl Rahner believes that matter has an intrinsic predisposition to evolve toward spirit, and that humans are the apex of this process of the universe's self-realization [40], [41]. Is it possible that the midwife proposal and Christian theology speak of the same idea when describing humankind as *the* vehicle through which matter ultimately becomes information/spirit?² Rather not.

While the two might bear some superficial resemblance, the Christian version differs in at least three significant ways. Eastern-Orthodox theologian Dumitru Stăniloae provides a compelling and illuminating account of this Christian idea of spiritualization of matter, which is worth quoting at large: "The world was created in order that man, with the aid of the supreme spirit, might raise the world up to a supreme spiritualization, and this to the end that human beings might encounter God within a world that had become fully spiritualized through their own union with God. The world is created as a field where, through the world, man'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 man were the only one freely working within the world, he could not lead the world to a complete spiritualization, that is, to his own full encounter with God through the world. God makes use of his free working within the world in order to help man, so that through man's free work both he and the world may be raised up to God and so that, in cooperation with man, God may lead the world toward that state wherein it serves as a means of perfect transparency between man and himself" [43].

The first major difference between the Christian perspective and midwife posthumanism consists of the presence versus the absence of God throughout the process of spiritualization of the universe. Christian thought is, of course, theistic. It affirms, as Stăniloae makes it clear, that humans by themselves could never lead the world to fulfill its full potential of spiritualization without divine collaboration. Unsurprisingly, the midwife hypothesis makes no explicit mention of a deity, but it should not be too easily labeled as atheistic. The kind of cosmic harmony that Kurzweil speaks about when describing a universe infused with intelligence can better be categorized as a form of pantheism, rather than atheism.

Secondly, an even more profound difference between the two visions concerns their teleology. In the Christian perspective, the spiritualization of matter is not a goal in itself. Rather, it is only relevant within the larger picture of the free relationship of love between God and humans. Stăniloae explicitly articulates that in Christian theology the universe is only valuable "with a view to the ultimate and total encounter" between creator and creature. The ultimate purpose of the

² This parallel was first pointed out in [42].

world is therefore to facilitate this encounter. But in order to fulfill this role, the world needs to achieve perfect transparency, hence the need for “supreme spiritualization.”

The midwife proposal, on the other hand, exhibits a rather unconvincing teleology. If the purpose of the universe is indeed informatization, as Lovelock conjectures, or the saturation with intelligence, as Kurzweil proposes, one could legitimately ask: why? Furthermore, it might be useful to pursue the midwife proposal to the absurdity of its final outcome. What would happen after the goal of informatization, complexity and order is physically achieved? Would the universe continue to exist forever in that state of perfect equilibrium, known in physics as the ‘Big Freeze’ [44], and synonymous to a heat death? Or would it then explode again into a new universe through another Big Bang, in which case the question of meaning and teleology would simply be reported to the end of the next cycle? Although we could certainly imagine higher-level informational beings inhabiting such a transfigured cosmos, the question of purpose still remains. Without an eternal God, infinitely generating new knowledge and meaning, it is hard to imagine what else could give purpose to such beings. Although this might be due to inherent limitations in our current imagination, it could also signal a weak and unsatisfactory teleology from the part of the midwife hypothesis.

Thirdly, the midwife proposal can rightly be suspected to arise from a certain dissatisfaction with the human condition, hence the need to replace humans with more advanced beings that will “inherit the earth” [45]. Theological anthropology might have a few problems with this. All the monotheistic religions share the intuition that there is something special about humans. In Christian anthropology, this intuition is encapsulated in the doctrine of the *imago Dei*: humans bear in them the image of God. Moreover, Christian faith is built around the testimony that God became human through the incarnation of Jesus Christ. These two, the *imago Dei* and the incarnation, strongly imply that humans, limited and imperfect as they might be, are in a way *enough*.

The issue of teleology is again of critical importance. From an evolutionary perspective, it can be said that the purpose of the universe is indeed to ‘awaken.’ But this awakening does not need to entail the saturation of the cosmos with intelligence, by being transformed into an unthinkable big supercomputer. Instead, the universe awakens by naturally giving birth to a conscious entity, which possesses all the mental and moral capabilities necessary to become a recipient of divine revelation and enter into a relational covenant with God. In this case, the awakening process would further be validated and fulfilled in the incarnation of the divine *Logos*, what Teilhard de Chardin calls “the Omega Point” of cosmic evolution [39, pp. 250-275]. Christopher Fisher summarizes this point very well: “In theological perspective, the appearance of personal subjective self-awareness and transcendentalism in human beings means that there is no need for another step in biological or cosmic evolution [...]: the process is complete (complete, in particular, in the incarnation itself), having reached the goal of opening material reality directly to conscious relationship with the Absolute” [46].

In patristic anthropology, and in particular in the writings of Maximus the Confessor, human beings are described as microcosms, miniature recapitulations of the entire cosmos [47], [48]. It is thus possible to affirm that with the emergence of the human person, the cosmos itself becomes conscious. Analogically, in the human response to God’s calling to relationship, the cosmos itself is brought to fulfillment and potentially transfigured, as in Stăniloae’s cosmology.

Christian anthropology acknowledges the limitations of human nature and the need to transcend them, but it suggests a radically different solution from the one advocated by the midwife proposal.

Humans are called to transcend their nature through the pursuit of deification, or *theosis*. Far from being “little more than the Christian’s alternative to human enhancement” [42, p. 340], *theosis* implies a radical transformation of human nature at its most profound level. According to Stăniloae’s definition, *theosis* is the “greatest possible union with God wherein the fullness of God is stamped upon the human being, yet without the human being thereby being dissolved into God” [43, p. 89]. *Theosis* suggests transcending human nature by downsizing oneself through God’s kenotic self-giving love, in contrast to the expansion of the self, entailed by Kurzweil’s vision [42, p. 330].

Finally, it is interesting to observe how the midwife proposal still struggles to find a place for human distinctiveness in its story. Even though AI will eventually “match and then vastly exceed the refinement and suppleness of what we regard as the best of human traits,” Kurzweil still struggles to find a feature that remains uniquely human: “There will be no distinction, post-Singularity, between human and machine or between physical or virtual reality. If you wonder what will remain *unequivocally human* in such a world, it’s simply this quality: ours is the species that *inherently seeks to extend its physical and mental reach beyond current limitations*” [11, p. 9, my emphases].

Kurzweil does not explain how humans are different in this respect from the animals. After all, isn’t this tendency to reach beyond the limits inherent to biological life, in general? In theological anthropology, this exocentricity of human nature is yet another mark of the *imago Dei*: we continuously strive, most of the time unconsciously, towards a destiny of fellowship with God in the *eschaton*, as beautifully described by Wolfhart Pannenberg [49]. But Kurzweil’s vision predictably lacks such context. What causes this supposedly uniquely human restlessness? And what is its *telos*?

The midwife proposal has no answers to these questions. While it might bear some superficial resemblance to Rahner’s openness to transcendence or with Pannenberg’s exocentricity, it does not come even close to painting as coherent a picture as these theological proposals.

For Rahner, humans are indeed intrinsically open to transcendence, but this is only because of the pre-apprehension of the infinite reality that is the transcendent God [40, p. 33]. Similarly, the exocentricity proposed by Pannenberg is a metaphysical drive toward fulfilling a vocational destiny in the encounter with God, who is the source of both the drive and of direction [50]. In both theological accounts, the typically-human longing for transcendence only makes sense if there exists an infinite transcendence, namely God, to long for in the first place. One might not agree with the inherent theological assumptions, but the system is at least self-coherent.

For Kurzweil, to the contrary, this thirst to exceed limitations is ultimately empty of content: it exists only because it is a necessary prerequisite for developing the kind of technology that can saturate the cosmos with intelligence. When compared to the theological accounts of human distinctiveness as *imago Dei*, the midwife proposal looks shallow and highly unconvincing.

VII. THE MIDWIFE PROPOSAL IS MORALLY DANGEROUS

While the theological and philosophical weakness of the midwife proposal might not be too imperative, its ethical ramifications are genuinely dangerous and in need of urgent clarification. Firstly, if saturating the universe with intelligence is the ultimate cosmic goal, then it follows that everything should be evaluated according to the measure in which it advances or hinders this process. This is already visible in how some in the bio-liberal movement are pleading that becoming posthuman is a moral imperative [51]. If such a view

becomes mainstream, how will the value of each individual human being come to be judged? Would human life be valuable in itself, or only insofar as it contributed to the progress towards the singularity or the Novacene?

Secondly, and more generally, if technological progress towards the so-called awakening of the cosmos is regarded as the ultimate goal of *everything*, then this view has the potential to substantially alter our current ethical definitions of good and evil. 'Good' would in such a case be any effort that promotes the technological singularity or the Novacene. Any resistance to the mainstream paradigm, such as refusing to augment oneself, could become synonymous with moral evil and sanctioned accordingly.³ This might sound dystopic, but it is a result of following the thread of Kurzweil and Lovelock's ideas to their logical conclusion. Neither of them intentionally proposes such a chilling moral system, but certain readings of the midwife hypothesis could nevertheless lead in that direction.

Thirdly, as discussed earlier, there is no reason to believe that super-intelligent cyborgs would necessarily be strong AI. It is equally likely, or perhaps even more likely, that they would turn out to be mindless automatons, superbly capable to outsmart us, but totally incapable of feeling or thinking anything. Could a world populated and radically transformed by such automatons be one we would deem as 'good'? Without a doubt, no. Even if human evolution is taken as a proof that the universe selects for intelligent life-forms, not any kind of intelligence gets positively selected. If the universe 'wants' indeed to 'awaken', it is *our* type of intelligence that it ultimately needs, one that is also accompanied by subjective experience and understanding. Otherwise, what would be the point?

Undoubtedly, even the most convinced believer of the midwife proposal would agree that the universe doesn't seek to be saturated with a mindless type of intelligence. To make any sense, the midwife proposal needs strong AI. There is thus a hidden built-in assumption that our cyborg descendants will be intelligent not in the way that current AI programs are, but that they will also be centers of selfhood and phenomenal experience, truly capable of thinking, understanding, and feeling. As of today, we are still completely in the dark regarding this possibility. We simply do not know whether machines could ever become conscious. Although AI has made significant progress toward replicating human intelligence 'on the outside', from what we know it has made *zero* progress toward acquiring consciousness, or an inside-ness.

Thus, even from a non-theistic utilitarian perspective, which judges things to be good or bad depending on how efficiently they promote the wellbeing of conscious agents, the midwife hypothesis is deeply problematic in a weak AI scenario. The hypothesis relies therefore on the possibility of strong AI, something that is often not made explicit enough in its manifestos.

VIII. CONCLUSION

This succinct overview of the midwife proposal and its assumptions enables some provisional conclusions. First and foremost, the character of the ideas behind it is highly speculative. Although they are presented as sound scientific truths, even the briefest of analyses

³ This is similar to the outcome of a thought experiment known as Roko's Basilisk, where a future omnipotent artificial superintelligence (ASI) decides to retroactively reward those who promoted its existence and punish those who did not (by resurrecting them through avatar reconstruction and then torturing those avatars eternally), in order to motivate us in the present to invest everything we have in the pursuit of ASI for fear of retribution. Although this scenario might sound anything from logically flawed to hilarious, it has caused a lot of anxiety among members of the LessWrong virtual community of rationalists [52]

reveals this to be an overstatement. This is also reflected by the skepticism with which the scientific and AI communities continue to regard such views [53].

That being said, Kurzweil's core idea that technological progress is accelerating, even though perhaps at a slower pace than he suggests, is still a valuable observation. Similarly, Lovelock's creative imagery of how the Novacene world will look like is very powerful: electronic animals grazing solar-powered plants, robots so small and fast that they inhabit and study the quantum world, or cyborgs thinking so fast that "the experience of watching your garden grow gives you some idea of how future AI systems will feel when observing human life" [12, p. 82]. However, any value of such images and ideas is outbalanced by their questionable science, unclear teleology and dangerous ethical implications.

The principles upon which the midwife proposal is based are far from being merely 'neutral' scientific and technical observations. As shown in the paper, they stem from a philosophically dubious understanding of the purpose of the universe and the role of humans. The idea that humanity has a seminal role to play in cosmic evolution by developing AI might be intriguing, but at a closer inspection it is exposed to be lacking support, depth, and a coherent teleology, especially in comparison with theological accounts of human distinctiveness. Finally, the *doctrine* that humanity has only a midwife role to play in the larger narrative of the cosmos evolving toward hyper-intelligence comes with heavy and rather indefensible ethical implications for the value of human life and the very definition of good and evil.

The midwife proposal therefore makes claims that, although couched in scientific language, belong more to the realm of religious discourse. Even when judged solely by their internal logic and coherence, such anthropologies and cosmologies fare much worse than their Christian counterparts, on which they draw. Before becoming too quickly resigned to a fate of collective demise, we should perhaps stop and wonder whether the future doesn't, in fact, badly need us.

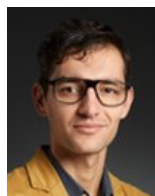
ACKNOWLEDGMENT

This publication was made possible through the support of a grant from the Templeton World Charity Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Templeton World Charity Foundation.

REFERENCES

- [1] B. Joy, "Why the future does not need us," *Wired*, 2001. Accessed Jan. 28, 2021. [Online]. Available: <https://www.wired.com/2000/04/joy-2/>.
- [2] S. Papert, "The summer vision project," *MIT AI Memos (1959 - 2004)*, 1966. Accessed Jan. 28, 2021. [Online]. Available: <http://people.csail.mit.edu/brooks/idocs/AIM-100.pdf>.
- [3] S. J. Russell and P. Norvig, *Artificial intelligence: A modern approach*, Upper Saddle River, NJ: Prentice Hall, 2003, p. 21.
- [4] H. Moravec, *Mind children: The future of robot and human intelligence*, Cambridge, MA: Harvard University Press, 1988, p. 15.
- [5] I.J. Good, "Speculations concerning the first ultraintelligent machine," *Advances in Computers*, vol. 6, pp. 31-88, 1965, doi: 10.1016/S0065-2458(08)60418-0.
- [6] N. Bostrom, *Superintelligence: Paths, dangers, strategies*, Oxford, UK: Oxford University Press, 2014.
- [7] J. Searle, "Minds, brains and programs," *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417-457, 1980, doi: 10.1017/S0140525X00005756.
- [8] D. Chalmers, "Facing up to the problem of consciousness," *Journal of Consciousness Studies*, vol. 2, no. 3, pp. 200-219, 1995.
- [9] V. C. Müller and N. Bostrom, "Future progress in artificial intelligence: A survey of expert opinion," in *Fundamental Issues of Artificial Intelligence*,

- V. C. Müller Ed. Berlin, DE: Springer, 2016, pp. 553-571.
- [10] M. Sparkes, "Top scientists call for caution over artificial intelligence," *The Telegraph*, 2015. Accessed Jan. 28, 2021. [Online]. Available: <https://www.telegraph.co.uk/technology/news/11342200/Top-scientists-call-for-caution-over-artificial-intelligence.html>.
- [11] R. Kurzweil, *The singularity is near: When humans transcend biology*, New York, NY: Viking Penguin, 2005.
- [12] J. Lovelock, *Novacene: The coming age of hyperintelligence*, London, UK: Allen Lane, 2019.
- [13] M. Shanahan, *The technological singularity*, Cambridge, MA: MIT Press, 2015, p. 233.
- [14] V. Vinge, "The coming technological singularity: How to survive in the post-human era," in *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, G. A. Landis Ed. NASA Publication CP-10129, 1993, pp. 11-22.
- [15] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, 1965.
- [16] R. Kurzweil, *The age of spiritual machines*, New York, NY: Viking Press, 1999, pp. 30-32.
- [17] J. Lovelock, "Gaia as seen through the atmosphere," *Atmospheric Environment*, vol. 6, no. 8, pp. 579-580, 1972, doi: 10.1016/0004-6981(72)90076-5.
- [18] M. Dorobantu, "Human-Level, but Non-Humanlike: Artificial Intelligence and a Multi-Level Relational Interpretation of the Imago Dei," *Philosophy, Theology and the Sciences*, forthcoming 2021.
- [19] D. Kahneman, *Thinking, fast and slow*, New York, NY: Farrar, Straus and Giroux, 2011.
- [20] P. J. Barnard and J.D. Teasdale, "Interacting cognitive subsystems: A systemic approach to cognitive-affective interaction and change," *Cognition and Emotion*, vol. 5, no. 1, pp. 1-39, 1991, doi: 10.1080/02699939108411021.
- [21] J. Haidt, *The happiness hypothesis: Finding modern truth in ancient wisdom*, New York, NY: Basic Books, 2006.
- [22] I. McGilchrist, *The master and his emissary: The divided brain and the making of the Western world*, Yale University Press, 2009.
- [23] H. L. Dreyfus and S. L. Dreyfus, *Mind over machine: The power of human intuition and expertise in the era of the computer*, Free Press, 1986, pp. 16-51.
- [24] B. Zhang, Q.-Y. Cai, M.-S. Zhan and L. You, "Information conservation is fundamental: Recovering the lost information in Hawking radiation," *International Journal of Modern Physics D*, vol. 22, no. 12, 1341014, 2013, doi: 10.1142/S0218271813410149.
- [25] Genesis 1: 27.
- [26] M. Cortez, *Theological anthropology: A guide for the perplexed*, London, UK: T&T Clark, 2010, pp. 21-24.
- [27] Genesis 1: 28.
- [28] B. Singler, "Existential Hope and Existential Despair in AI Apocalypticism and Transhumanism," *Zygon: Journal of Religion & Science*, vol. 54, no. 1, pp. 156-176, 2019, doi: 10.1111/zygo.12494.
- [29] R. M. Geraci, "Apocalyptic AI: Religion and the promise of artificial intelligence," *Journal of the American Academy of Religion*, vol.76, no. 1, pp. 138-166, 2008, doi: 10.1093/jaarel/lfm101, p. 158.
- [30] R. Cole-Turner, "The Singularity and the rapture: Transhumanist and popular Christian views of the future," *Zygon: Journal of Religion & Science*, vol. 47, no. 4, pp. 777-796, 2012, doi: 10.1111/j.1467-9744.2012.01293.x.
- [31] Romans 4: 13.
- [32] W. Grassie, "Politics by other means: Science and religion in the twenty-first century," Bryn-Mawr, PA: Metanexus, 2010, p. 282.
- [33] G. Dvorsky, "Stephen Hawking says A.I. could be our 'worst mistake in history,'" *Gizmodo*, 2014. Accessed Jan. 28, 2021. [Online]. Available: <https://io9.gizmodo.com/stephen-hawking-says-a-i-could-be-our-worst-mistake-in-1570963874>.
- [34] E. M. Jones, "'Where is everybody?': An account of Fermi's question," *Los Alamos National Laboratory*, USA, 1985, doi: 10.2172/5746675. Accessed Jan. 28, 2021. [Online]. Available: <https://www.osti.gov/servlets/purl/5746675>.
- [35] Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters," *Astronomy & Astrophysics*, vol. 594, article no. A13, 2016, doi: 10.1051/0004-6361/201525830, p. 32.
- [36] A. Sandberg, E. Drexler, and T. Ord, "Dissolving the Fermi paradox," *ArXiv*, 2018. Accessed Jan. 28, 2021. [Online]. Available: <https://arxiv.org/abs/1806.02404>.
- [37] G. Kaufman, "The concept of nature: A problem for theology," *Harvard Theological Review*, vol. 65, no. 3, pp. 337-366, 1972, doi: <https://doi.org/10.1017/S0017816000001619>.
- [38] K. Junghyung, "Christian anthropology in an age of science: Between anthropocentrism and non-anthropocentrism," *The Expository Times*, vol. 129, no. 12, pp. 547-553, 2018, doi: 10.1177/0014524617753327.
- [39] P. Teilhard de Chardin, *The phenomenon of man*, New York, NY: Harper & Row, 1961, p. 220.
- [40] K. Rahner, *Foundations of Christian faith: An introduction to the idea of Christianity*, translated by W. V. Dych, New York, NY: Crossroad, 1982.
- [41] O. Putz, "Evolutionary biology in the theology of Karl Rahner," *Philosophy and Theology*, vol. 17, no. 1&2, pp. 85-105, 2005, doi: 10.5840/philtheol2005171/25.
- [42] R. Cole-Turner, "Theosis and human enhancement," *Theology and Science*, vol. 16, no. 3, pp. 330-342, 2018, doi: 10.1080/14746700.2018.1488526, p. 336.
- [43] D. Stăniloae, *The experience of God: Orthodox dogmatic theology. Vol. 2 – The world: Creation and deification*, translated by I. Ionita and R. Baringer, Brookline, MA: Holy Cross Press, 2000, p. 59.
- [44] A. V. Yurov, A. V. Astashenok, and P. F. Gonzalez-Diaz, "Astronomical bounds on a future Big Freeze singularity," *Gravitation and Cosmology* vol. 14, pp. 205-212, 2008, doi: 10.1134/S0202289308030018.
- [45] Matthew 5: 5.
- [46] C. L. Fisher, "Animals, humans and x-men: Human uniqueness and the meaning of personhood," *Theology and Science*, vol. 3, no. 3, pp. 291-314, 2005, doi: 10.1080/14746700500317289, p. 307.
- [47] Maximus the Confessor, "Epistolae VI," in *Patrologiae Cursus Completus*, J.-P. Migne Ed. Tomus 91, Paris, FR: Garnier Freres, 1865, col. 429.
- [48] L. Thunberg, *Microcosm and mediator: The theological anthropology of Maximus the Confessor* (2nd edition), Chicago and La Salle, IL: Open Court, 1995, pp. 132-143.
- [49] F. LeRon Shults, *Reforming theological anthropology: After the philosophical turn to relationality*, Grand Rapids, MI: Eerdmans, 2003, p. 235.
- [50] W. Pannenberg, *What is man? Contemporary anthropology in theological perspective*, Philadelphia, PA: Fortress Press, 1970, p. 220.
- [51] M. Walker, "Ship of fools: Why transhumanism is the best bet to prevent the extinction of civilization," in *H±: Transhumanism and Its Critics*, G. R. Hansell and W. Grassie Eds. Philadelphia, PA: Metanexus Institute, pp. 94-111, 2011, p. 95.
- [52] B. Singler, "Roko's Basilisk or Pascal's? Thinking of Singularity Thought Experiments as Implicit Religion," vol. 20, no. 3, pp. 279-297, 2018, doi: 10.1558/imre.35900.
- [53] J. Rennie, "Ray Kurzweil's slippery futurism," *IEEE Spectrum*, 2010. Accessed Jan. 28, 2021. [Online]. Available: <https://spectrum.ieee.org/computing/software/ray-kurzweils-slippery-futurism>.



Marius Dorobantu

Dr. Marius Dorobantu is a research associate and lecturer at the Vrije Universiteit Amsterdam, working on issues of science & religion. His current project, financed with a "Diverse Intelligences" grant from the Templeton World Charity Foundation, is entitled "Understanding Spiritual Intelligence: Computational, Psychological and Theological Approaches." Formerly he obtained a PhD in ethics from the University of Strasbourg, France in 2020, with a thesis entitled "Theological Anthropology and the Possibility of Human-Level Artificial Intelligence: Rethinking Human Distinctiveness and the Imago Dei."