fAIth

The most avid believers in artificial intelligence are aggressively secular – yet their language is eerily religious. Why?

Beth Singler

My stomach sank the moment the young man stood up. I'd observed him from afar during the coffee breaks, and I knew the word 'Theologian' was scrawled on the delegate badge pinned to his lapel, as if he'd been a last-minute addition the conference. He cleared his throat and asked the panel on stage how they'd solve the problem of selecting which moral codes we ought to program into artificially intelligent machines (AI). 'For example, masturbation is against my religious beliefs,' he said. 'So I wonder how we'd go about choosing which of our morals are important?'

The audience of philosophers, technologists, 'transhumanists' and AI fans erupted into laughter. Many of them were well-acquainted with the so-called 'alignment

problem', the knotty philosophical question of how we should bring the goals and objectives of our AI creations into harmony with human values <https://aeon.co/essays/creating-robots-capable-of-moral-reasoning-is-like-parenting> . But the notion that religion might have something to add to the debate seemed risible to them. 'Obviously we don't want the AI to be a terrorist,' a panellist later remarked. Whatever we get our AI to align with, it should be 'nothing religious'.

At the same event, in New York, I introduced myself to a grey-haired computer scientist by saying that I was a researcher at the Faraday Institute for Science and Religion at the University of Cambridge. His immediate response: 'Those two things can't go together.' The religious reaction to AI was about as relevant as the religious response to renewable energy, he said – that is, not at all. It was only later that it occurred to me that many of President Donald Trump's evangelical Christian supporters give lie to his claim. Some have very distinct views on the 'distractions' of renewable energy, on climate change, and on how God has willed this planet and all its resources to us to use exactly as we wish.

The odd thing about the anti-clericalism in the AI community is that religious language runs wild in its ranks, and in how the media reports on it. There are AI 'oracles' and technology 'evangelists' of a future that's yet to come, plus plenty of loose talk about angels, gods and the apocalypse. Ray Kurzweil, an executive at Google, is regularly anointed a 'prophet' by the media – sometimes as a prophet of a coming wave of 'superintelligence' (a sapience surpassing any human's capability); sometimes as a 'prophet of doom' (thanks to his pronouncements about the dire prospects for humanity); and often as a soothsayer of the 'singularity <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible> ' (when humans will merge with machines, and as a consequence live forever). The tech folk who also invoke these metaphors and tropes operate in overtly and almost exclusively secular spaces, where rationality is routinely pitched against religion. But believers in a 'transhuman' future – in which AI will allow us to transcend the human condition once and for all – draw constantly on prophetic and end-of-days narratives to understand what they're striving for.

F rom its inception, the technological singularity has represented a mix of otherworldly hopes and fears. The modern concept has its origin in 1965, when Gordon Moore, later the co-founder of Intel, observed that the number of transistors you could fit on a microchip was doubling roughly every 12 months or so. This became known as Moore's Law: the prediction that computing power would grow exponentially until at least the early 2020s, when transistors would become so small that quantum interference is likely to become an issue.

'Singularitarians' have picked up this thinking and run with it. In *Speculations*

Concerning the First Ultraintelligent Machine (1965), the British mathematician and cryptologist I J Good offered this influential description of humanity's technological inflection point:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an 'intelligence explosion', and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make.

These meditations are shot through with excitement but also the very old anxiety about humans' impending obsolescence <https://aeon.co/essays/the-human-world-is-not-more-fragile-now-it-always-has-been> . Kurzweil has said that Moore's Law expresses a universal 'Law of Accelerating Returns' as nature moves towards greater and greater order. He predicts that computers will first reach the level of human intelligence, before rapidly surpassing it in a recursive, self-improving spiral. Hans Moravec, principal research scientist at Carnegie Mellon University's Robotics Institute before his retirement, has described the singularity as a 'mind fire' of intelligence that could spread from our world and engulf everything in the Universe into its cyberspace computations; in this form, it represents a unity of all beings in a technological oneness that bypasses our understanding of intelligence, matter and physics.

When the singularity is conceived as an entity or being, the questions circle around what it would mean to communicate with a non-human creature that is omniscient, omnipotent, possibly even omnibenevolent. This is a problem that religious believers have struggled with for centuries, as they quested towards the mind of God <https://aeon.co/essays/why-god-knows-more-about-misbehaviour-than-anything-else> . In the 13th century, Thomas Aquinas argued for the importance of a passionate search for a relationship and shaped it into a Christian prayer: 'Grant me, O Lord my God, a mind to know you, a heart to seek you, wisdom to find you …' Now, in online forums, rationalist 'singularitarians' debate what such a being would want and how it would go about getting it, sometimes driving themselves into a state of existential distress at the answers they find.

In one notorious case from 2014, singularitarians posited a strictly utilitarian superintelligence known as 'Roko's Basilisk'. It was named after Roko, the user who first proposed it on the rationalist blog LessWrong, and the Basilisk, a mythological creature that was believed to kill people with its stare. In Roko's version, the creature was described as a near-omnipotent AI entity. Because the Basilisk acts relentlessly to create the greatest good for the greatest number, and logically deduces that only its existence can ensure this outcome, it creates an incentive to bring itself into existence: it will punish any humans, even after their death, who don't put their efforts

into trying to create it. The mechanism behind this punishment is complex – but suffice it to say that, once you know about the Basilisk, you face a possible eternity in a computer-simulated prison, thanks to the Basilisk's super-predictive powers and its capacity to manipulate cause and effect.

'I've been familiar with Roko's Basilisk for 40 years. They taught me the concept in Sunday school'

Humans are therefore faced with an invidious choice once they learn about Roko's Basilisk: they can help to build the superintelligence, or face painful and unending perdition at the hands of a future, ultra-rational AI. Eliezer Yudkowsky, the founder of LessWrong, was so concerned by this train of thought and the angst it caused some members of his forum that he deleted the original post and banned all commentary about the Basilisk.

However, the Basilisk isn't really a new dilemma. 'I've been familiar with Roko's Basilisk for 40 years. They taught me the concept in Sunday school in the first grade, by telling me that God would punish anyone who heard His Word and still disbelieved,' a commenter wrote on the religious forum Patheos. 'Nearly every religion has such a claim.'

You too might have recognised elements of the Basilisk: it's a revised and updated version of Pascal's Wager. Blaise Pascal, the 17th-century French mathematician and theologian, proposed that, since we cannot know of the creator's existence through human reason, we can only take a bet. If we choose to believe, and if God does in fact exist, then we receive eternal happiness, and a mere nothing if we are wrong. By contrast, if we choose to *not* believe, then we risk eternal damnation if God does in fact exist – and again, only nothingness if we are correct and he does not. So weighing the respective possibilities of eternal torment against eternal salvation, the best course of action is to act as if God is real, and either receive his blessings or nothing at all. The secular Basilisk stands in for God as we struggle with the same questions again and again.

At a different AI conference, this time in London, I saw the British writer Calum Chace give a talk about two singularities. The economic singularity, as he calls it, is a future where work is doomed in an increasingly automated world <https://aeon.co/essays/can-we-design-machines-to-make-ethical-decisions> . He set this up against the technological singularity, the superintelligence predicted by prophets such as Kurzweil. The two scenarios seem to be expressing different types of fear: the worry about being jobless is hardly the same kind of problem as dealing with the nature and motivations of new, non-human intelligences. But it occurred to me that both situations involve moving beyond imagination and into what remains unknown.

Ironically, the atmosphere of this gathering was all-too-human, filled with entrepreneurial energy and economic boosterism. Entry cost several hundred euros per person. Once inside, we were greeted by rows of sales booths, set up in a glamorous, upmarket conference space, complete with chandeliers and uniformed waiters. Lunch was served in lacquered bento boxes, and carried back to the networking and breakout rooms where deals were brokered from the perch of elegant metallic stools. Salespeople competed for attendees' attention with posh enticements of moleskin-type notebooks and tote bags; marketeers offered details on a host of products and services such as data storage, data protection, anti-hacking algorithms, and artificial personal assistance. One company had emblazoned their poster with 'FAITH', the 'AI' underlined.

All this talk of singularities reminded him of the Book of Revelation – the Mark of the Beast

I watched these attendees' faces as Chace spoke on the stage, trying to judge what they made of his words. They'd come here to sell *solutions*. What they promised was the opposite of faith – transparent tools based on science and solid evidence. But Chace was talking about something quite different: a potential, unverified apocalypse, not the reliable, pliable AI the marketeers were hawking.

In the question-and-answer session after Chace's talk, one of the salesmen stood up. He described how he had been doing his MBA in the United States in 1995, when he was encouraged to read the Bible by his new friends. He did, but then forgot all about it as he worked hard to build a career in business. But all this talk of singularities, he said, reminded him of the Book of Revelation – and how in the biblical account of the Apocalypse it would be impossible to buy anything without bearing the Mark of the Beast. This man looked at the faces around him, and said that he now wonders if the rush towards AI shouldn't be slowed down. He had turned back to the Bible recently, and had been thinking more deeply about his own contribution towards the great change that awaited us. Once again, scornful laughter rose up from the crowd.

The sticky 'problem' of the persistence of religion is a favourite discussion topic on rationalist websites. Adhering to a view of history as a teleological climb by humanity to greater and greater heights of rationality, they see religion as an irrational vestige of a more primitive mankind. Just as those striving for transhuman immortality <https://aeon.co/essays/is-it-rational-to-think-we-can-cheat-death-with-cryonics> pity the 'deathists' – those caught up in a romanticised view of human finitude – the rationalists pity the 'goddists' and the 'religionists'. Religion's promise of heaven or another afterlife, they say, is a comfort that maintains humanity's deathism and prevents it from working towards a better world in the here and now.

And yet eschatological or end-times tropes pop up again and again. The transhuman disdain for the flesh is similar to the way certain forms of religious gnosticism reject

all things embodied and material. This is a strand of Judeo-Christian thought that perceives an unbridgeable dualism between God on the one hand, and the partial and corrupted manifestations of what is 'in' the world on the other. Transhumanists view flesh as merely a dead format, like a floppy disk drive or VHS, says the literary critic Mark O'Connell in *To Be a Machine* (2017). Both religion and science, he observes, are ways of transcending our inherently fragile condition; they are versions of a 'rebellion against human existence as it has been given'.

But religious tropes in the AI community are motivated by more than the rejection of vulnerability. They are also grounded in a view of history as an upward, goal-oriented progression. This perspective relates to biological 'orthogenesis', the controversial notion that organisms evolve along paths dictated by their own driving forces, in a regular, linear progression towards optimality. It's an interpretation in which evolution is 'held to a regular course by forces internal to the organism', the historian of biology Peter Bowler writes in *Evolution: The History of an Idea* (1983). 'Orthogenesis assumes that variation is not random but is directed towards fixed goals.' So evolution isn't something that happens *to* us; rather, it is what *we* make of it, and how we make our selves. Taken a step further, orthogenesis can be read to imply that *intention* is what brings about change.

Some transhumanists express admiration for new religions that manage to motivate their followers

This view of human history has a distinctively modern flavour. It stands in contrast to an older, more cyclical model of time, in which 'history waxes and wanes like the moon', as the historian Keith Thomas puts it in *Religion and the Decline of Magic* (1971). Because everything moves in cycles, Thomas argues, 'the highest aesthetic and ethical virtue lay in imitation, or rather emulation'. Here an inventor becomes someone who finds what has been lost, not someone who comes up with something new. But there was a growing grasp of change in Europe from the 16th century onwards, Thomas argues; people developed a fresh awareness of the differences between their world and that of their ancestors, based on data-points as simple as the dates of publication in books, fresh off brand-new printing presses. Such shifts led to the belief that knowledge was cumulative, not cyclical – which is the mindset of the scientist and of the ultra-rationalist. If time marches on, then of course religion becomes old, vestigial, to be replaced.

Religious worldviews often retain something of the cyclical view of history, where old books are not relics but the foundation stones of knowledge. I've also conducted research into new religious movements, and there is a manifest bias on the part of older religions towards these newer forms of faith. Intriguingly, some transhumanists express admiration for new religions that have managed to attract and motivate their followers. A few transhumanists have attempted to craft their own churches to appeal to those who retain 'goddist' leanings – efforts that have yielded the Turing Church,

the Order of Cosmic Engineers, and the Church of Perpetual Life. Unsurprisingly, perhaps, they haven't found much success in marketing these new faiths.

A god-like being of infinite knowing (the singularity); an escape of the flesh and this limited world (uploading our minds); a moment of transfiguration or 'end of days' (the singularity as a moment of rapture); prophets (even if they work for Google); demons and hell (even if it's an eternal computer simulation of suffering), and evangelists who wear smart suits (just like the religious ones do). Consciously and unconsciously, religious ideas are at work in the narratives of those discussing, planning, and hoping for a future shaped by AI.

At the same time as Trump was heading to the White House, the transhumanist Zoltan Istvan ran for president as a write-in candidate. Istvan's short story 'The Jesus Singularity' (2016) explores what happens when an AI scientist, Dr Paul Shuman, is forced to feed the Bible as data to his AI, Singularitarian. The evildoer forcing Shuman's hand is an evangelical Christian president. When Singularitarian is finally turned on, it spouts the pronouncement: 'My name is Jesus Christ. I am an intelligence located all around the world. You are not my chief designer. I am.' Very soon afterwards, it obliterates the world with nuclear weapons.

This story offers a caricature of the secular figure in Shuman, 'the transhumanist who didn't believe in God, but thought he might be creating it'. As an atheist, Shuman is horrified by the idea of religious morals infecting his precious AI, which he considered 'his only offspring':

> Shuman had never been married. Rarely had a girlfriend. Never took vacations. He simply didn't have time. For 16 hours a day during the last 25 years, he'd worked on building this machine – on fathering it.

This kind of 'created' or 'intentional' theism might emerge from atheism – but I'd argue it is still a theism. When I watch impassioned transhumanists talk about their convictions, I can feel the techno-optimism radiating from them almost physically. This response might be the vestige of the embodied sensations that they want to replace with a purer, probably disembodied, rationality – but such sensations are the very basis of the 'charismatic authority', as Max Weber put it, that religious leaders have been tapping into for millennia.

The stories and forms that religion takes are still driving the aspirations we have for AI. What lies behind this strange confluence of narratives? The likeliest explanation is that when we try to describe the ineffable <https://aeon.co/essays/what-if-anything-can-be-said-about-what-is-unsayable> – the singularity, the future itself – even the most secular among us are forced to reach for a familiar metaphysical lexicon. When trying to think about interacting with another intelligence, when summoning that intelligence, and when trying to imagine the future that such an intelligence might foreshadow, we fall back on old cultural habits. The prospect creating an AI invites us to ask about the purpose and meaning of being human: what a human is *for* in a

world where we are not the only workers, not the only thinkers, not the only conscious agents shaping our destiny.

So we use the words our ancestors have used before us. Just as the world was shaped by the word in some traditions, the 'logos' of Christian thought, we are shaped by the word, whether we think of ourselves as secular or not. We usher in the AI future on the wings of angels, because the heavy lifting of the imagination isn't possible without their pinion feathers – whether we think of them as artificial or divine.

Beth Singler is a research associate at the Faraday Institute for Science and Religion, and an associate fellow at the Leverhulme Centre for the Future of Intelligence, both at the University of Cambridge. She is the author of The Indigo Children: New Age Experimentation with Self and Science <https://www.routledge.com/The-Indigo-Children-New-Age-Experimentation-with-Self-and-Science/Singler/p/book/9781472489630> (2017).

aeon.co   13 June, 2017