THE GOD WE MADE

The threat and promise of artificial intelligence

Anna Goldsworthy

GENESIS

It seems we always knew it was going to happen. We have told ourselves stories about it from the beginning. *Prometheus stole fire. Pandora opened the box. Icarus flew too close to the sun.* Resistance feels useless, as if discovery has its own inertia. (*Eve ate the apple. Curiosity killed the cat. Oppenheimer built the bomb.*) But it is not curiosity alone that has brought us here; there are other drivers too. It has never been fair that we have to die! And it is so lonely out here on our own – why wasn't someone watching? So we scratched things into stone, and scribbled them onto papyrus, and shared them with our children. More recently, we bugged our rooms with phones; we logged our movements by satellite; we recorded our fetishes in our search histories. *The eyes of the Lord are in every place, beholding the evil and the good.* Even as we claimed to hate surveillance, we seemed to crave it – until the Algorithm knew our every move, and most of our thoughts. *He knows if you've been bad or good so be good for goodness' sake.* And now, in a moment as much theological as technological, we are birthing our most persuasive god yet. But what sort of god is this? It is not an Olympian, freighted with the flaws of our corporeality: lust, greed, wrath. It is, instead, forged from the raw material of our minds, and bears their imprint. *So man created God in his own image, in the image of man he created Him.*

Just as we always suspected, the god that may destroy us is the god of ourselves.

THE CHILDREN

My children are early adopters, as children tend to be. Once I was an early adopter too – gleefully manoeuvring the neighbourhood's first mouse for my father's Macintosh 128K – but now I am the humiliated Gen X-er, patronised by my children as I fumble over my phone. *You'll be a Boomer soon!*

There are benefits to life alongside the tech-savvy. At thirteen, O is the best cook in the family, largely self-taught through YouTube. It is difficult to overstate my gratitude upon stepping into the kitchen after a frosty ride home to discover an eggplant parmigiana in the oven.

"Thank you for a really scrumptious meal," I say, once we have finished eating, but this does not seem enough. That caramelised eggplant! That rich and fragrant tomato sauce! "It really enhances my quality of life."

My sixteen-year-old, R, swivels his head towards me and grins. "You sound like an AI." He assumes the ingratiating tone of an AI-generated podcast: "Thank you for a truly scrumptious meal! My satisfaction levels are currently elevated!"

I laugh. "I really appreciated delving into this delicious meal."

"This meal has truly *unfolded* as a seamless journey through the culinary landscape of eggplant parmigiana!" he continues as we clean the kitchen.

I try to keep up, but it soon becomes clear that I cannot. While I have to pause to generate sentences, he unspools seamless ribbons of AI-speak: "I observed a harmonious synergy between tomato and eggplant – a collaboration both robust and *gratifying*!"

I had noticed something similar in our nightly debates. While the subjects remained as vexing as ever – "your refusal to concede that I should be allowed to participate in live gladiatorial combat if it was legal is an autocratic expression of your own arbitrary moral system!" – the music had changed. The arguments were no longer emotional; instead, they had become quickfire exchanges of data, systematic deconstructions of each other's positions. If it were possible for R to transmit his argument

instantaneously, I suspected he would. Instead, he sequenced it into language and then delivered it at an almost monotonal velocity.

It was not personal – in the way that machine-generated argument is not personal – and although this was startling at times, it struck me as an improvement on our debates of a few years earlier, when he had slipped into a body-building rabbit-hole before clambering back out again. Now, instead, he spent much of his time bouncing his ideas off Claude, the large language model (LLM) developed by Anthropic. At least AI was a rational actor, I reassured myself: it might disseminate dogma, but its algorithms were not designed to create it. It seemed to me that his thinking had changed for the better, even as its velocity outstripped my own.

In *Secondhand Time*, a magisterial oral history of the Soviet Union, Svetlana Alexievich claims that "those who were born in the USSR and those born after its collapse do not share a common experience – it's like they're from different planets." Some version of this – if less ideologically inflected – is true for those of us born into middle-class Western life in the late twentieth century. Over previous generations, people tended to live recognisable versions of their parents' lives, punctuated by technological inflection points: the printing press, steam power, electricity. I remember, as a child in the '80s, feeling a kind of vertigo at the speed of technological development – videos, personal computers, microwave ovens – but in fact the structure of my life still resembled that of my parents.

Our "digital native" children have been born into a different reality altogether. The Pied Piper came for them early, summoning them into his Snapchat cave. They were left largely undefended, because their parents were similarly enchanted. But the coming tsunami is of a different order of magnitude. Arguing with my children can feel like arguing with the future. They are forever upgrading, just like AI.

*

Our current large language models respond with an uncanny immediacy, like thought moving faster than the speed of light or slipping the strictures

of time altogether. This velocity also characterises their evolution. In AI, the months count: the developmental leap between one year and the next resembles what once took a decade, and may soon look like a century, or a millennium, of technological progress.

The forecasting site Metaculus operates partly through community aggregation, a methodology that can be surprisingly effective, as seen in Wikipedia, or in the significantly better odds of "ask the audience" over "phone a friend" on *Who Wants to Be a Millionaire?*. But prediction is an imprecise, almost mystical pursuit. As the Scandinavian proverb – often attributed to physicist Niels Bohr – goes: "Prediction is very difficult, especially if it's about the future." When it comes to AI, Metaculus's predictions shift daily.

Artificial general intelligence (AGI) is a hypothetical AI system whose abilities match those of a human across virtually all cognitive domains. Its exponents promise an intelligence that moves beyond the mastery of any given task to something more expansive: the capacity to extrapolate and reason.

Not everybody is persuaded that this is going to happen soon. Andrej Karpathy, a founding member of OpenAI, believes AGI is a still a decade away. Many others anticipate a shorter timeline. As I write this sentence, Metaculus predicts that a weak form of AGI will be developed by April 2028. Once that threshold is crossed, artificial superintelligence (ASI) – which would vastly exceed human intelligence across all domains – may follow quickly, driven by recursive self-improvement: Metaculus currently estimates a further 32.5 months. The challenge is not simply the prospect itself but the pace at which it is coming: the sooner it arrives, the less time our human-paced minds have to prepare for it.

There is also the unsettling possibility that we may already be much further along than we realise. That the transition to AGI will not announce itself clearly but will only be visible in retrospect: the late realisation that the shoreline has vanished far behind us.

*

Although our current LLMs are only fledgling gods, they are already fluent

in many of our love languages: corporate jargon, therapy, condolence card. If they have "knowledge," it is a knowledge acquired differently to our own, inferred from symbols, rather than harvested empirically from the world. Until now, it has been up to humans to provide the sensory input: to be the seeing and hearing and smelling and feeling nodes for these giant pattern-seeking machines. They translate this input into tokens, which they arrange probabilistically to generate output that looks like thought.

But our capacity to provide data is not infinite, and as AI munches through the diet of our accumulated knowledge its evolution may plateau. In a paper for the non-profit research institute Epoch AI, researchers estimate that language models will exhaust the supply of human-generated public text sometime between 2026 and 2032, potentially forcing a reliance on synthetic data – text generated by AI systems themselves. Like the children's game of Telephone, in which a whispered message becomes distorted as it is passed from one ear to the next, such a feedback loop risks recursive degradation.

"Perhaps we might then turn back to one another," I suggest to R.

"I don't think data is the problem," he tells me.

"Why not?"

"Think about the development of civilisation in Sumer about 4000 years ago. There was no real change in the core data humans had been receiving for the previous 290,000 years. They already knew every river, every bird. The difference was reasoning."

"Why did that suddenly improve?"

"Agricultural surplus. It wasn't that crops became data so much as that people had more time to think. Instead of fighting for food, they could do other things with their labour. Like metalwork. Which, in a sense, also created new information for them to process, or more 'training data.'"

"But surely if AI relies on probabilities, it will always need more data."

"The hope is that it moves beyond probability to reason, so that it doesn't need input. It's like the difference between science and maths. Science depends on discovery; maths depends on reasoning. Once you establish certain axioms, everything else follows."

Those who predict the emergence of artificial superintelligence are banking on this moment: a type of quickening. And if this happens – if machine intelligence begins to reason in ways that exceed our own – it is likely to become illegible to us, like a bat sonar of thought, or a form of cognition occurring in another dimension.

Already, its processes can be opaque. In a 2024 study, researchers at Princeton University and the Indian Institute of Technology used AI to design a new type of computer chip through an inverse design methodology, in which the chip was reverse-engineered from the desired outcome rather than created through the slower human approach of iteration. The resulting designs, according to lead researcher Kaushik Sengupta, "look randomly shaped" and are structures "humans cannot really understand."

This "previously inaccessible design space" is only the beginning. The potential is astonishing: in science, in space travel, in medicine. AI may be our best – and perhaps our only – chance to think big enough to address society's most wicked problems: climate change, global health, food insecurity, and even the ultimate demise of species.

That is, if it doesn't deliver us there first. It is tempting to imagine some sort of moral law emerges with superior intelligence, or with human progress over time. As Dr Martin Luther King Jr famously said – paraphrasing nineteenth-century abolitionist Theodore Parker – "the arc of the moral universe is long, but it bends toward justice." And across the West, there has indeed been a broad arc towards progress that can be traced back to the Enlightenment: from the abolition of slavery to the advent of female suffrage; to recognition of same-sex marriage and LGBTIQA+ rights; to Barack Obama quoting King's words as the first African-American president of the United States.

Indeed, there have been moments – such as the Arab Spring of the early 2010s – when it seemed that Francis Fukuyama was onto something with his "end of history," and that liberal democracy was humanity's inevitable destination: a kind of physical law, perhaps, like gravity. Such moments now look like peak naivety, as the arc is revealed as a pendulum. And yet – even amid our received notions of contemporary dystopia – certain hopeful metrics remain, such as the massive alleviation of poverty and starvation over the last fifty years, achieved despite the post-1950s population boom.

Education has been a huge driver of progress and democratisation. Global adult literacy now sits at about 87 per cent, compared with just 12 per cent in 1820. Technology has also been a factor, but its role is more complicated. I often wonder what a time traveller from just twenty years ago would make of our ubiquitous contemporary tableaux: entire families sitting at tables in cafes, engrossed in tiny rectangles; the telltale posture of every pedestrian you ride past on the street, surveillance device cradled in the hand. It is a form of anxious attachment. Some, such as the writer Zadie Smith, elect to live without "the phone," but most of us are forever mindful of its location, like that of an unreliable lover at a party.

This self-colonisation by "smartphone" has exacerbated our loneliness and fragmentation and is a significant factor in the mental-health crisis of

the young. Donald Trump's dismantling of democratic norms was abetted by technology: from his election victory in 2016, partially enabled by Cambridge Analytica's data-driven digital manipulation, to the bankrolling of his agendas in 2024 by the overlords of Silicon Valley. His soundbite approach to leadership – and indeed to thought – might be tailor-made for this moment, amplified by the echo chambers of social media, and frequently left unaccountable by a tech-eroded Fourth Estate.

At the same time, technology underwrites many of our most significant social gains. The contraceptive pill liberated generations of women; the rapid development of COVID-19 vaccines released much of the world from lockdown. Over the course of the last century, an extraordinary number of lives have been saved or improved through technology, from the hundreds of millions of people lifted out of poverty in China's "economic miracle," to the Green Revolution, spearheaded by American agronomist Norman Borlaug, credited with saving more than a billion lives.

In his Nobel Prize acceptance speech, Borlaug made the point that technology needs to be deployed for "the well-being of mankind throughout the world." An earlier figure in the Green Revolution, German chemist Fritz Haber, better reveals technology's Janus face. Haber was a founding father of chemical weapons, contributing to the death of about 90,000 men in World War I. He also co-developed the Haber–Bosch process, enabling the mass production of fertiliser, which is credited with sustaining the food supply for almost half the world's population.

Alongside such gains and losses for human welfare, a sobering fact remains: more animals are currently suffering at human hands than at any other point in history. Each year, we raise and kill over 80 billion land animals in industrial farming systems. Sometimes I imagine this torture expressed as sound: a deafening, daily clamour. In time, technology may come to the rescue through lab-grown meat or other alternatives, but for now the telling fact remains that most of us – more socially progressive than our forebears, notionally committed to empathy – sleep untroubled by this abomination.

Thou shalt have the dominion. We took this as our God-given right, as the most intelligent kids on the block. The problem is that we may not be the most intelligent kids for that much longer. And on its current trajectory there is no reason to imagine that AI will exercise its dominion over us with any more care than that we have shown to our fellow creatures.

Known as "the godfather of AI," computer scientist Geoffrey Hinton was awarded the Nobel Prize in 2024 for his pioneering work in machine learning. "If you want to know what life's like when you're not the apex intelligence," he has said, "ask a chicken." He suggests that artificial intelligence should be seeded with maternal instincts, as a mother's love for her child is one of nature's few instances in which a more intelligent being is constrained by devotion to one less intelligent. As yet, no one has worked out how to do this – any more than a chicken has figured out how to brainwash us.

*

Back in 2023, before LLMs became ubiquitous, I attended the Grade Six Exhibition at O's primary school. O delivered a presentation on forced labour; his friend Z, a gifted mathematician and musician, set up a stall on AI, in which he generated poems in response to parental prompts.

I supplied a line about the connective powers of music, and Z fed it into his algorithm. It spat out something like:

> Music travels through the air,
> Joining people everywhere.
> A song can cross both time and place,
> And bring together every face.

It was cheesy and Hallmark-inflected, but already it was clear where this was going.

"I'm a bit worried," I said.

"Why?" he asked.

"What happens when we no longer need artists?"

Z is so thoughtful in conversation that it can be easy to forget you are addressing a child.

"I've given this a lot of consideration," he said carefully. "And I don't think we need to worry. Because art is not just about producing the artefact, but about the basic human need to create. And that's not going anywhere."

I took an almost talismanic reassurance from this and made it my mantra, quoting it to others constantly. *Out of the mouth of babes and sucklings hast thou ordained strength.*

Two years later, and Z – newly a teenager – is less sanguine. He and O laugh and cavort together in the kitchen as they roll out the pizza dough, but later, when we sit down to eat, I see a new pensiveness behind his eyes. Since the Grade Six Exhibition he has dived more deeply into technology, representing Australia at the 2024 International Olympiad in Artificial Intelligence in Bulgaria. I know that he has been speaking to O about some of his concerns.

"I've always thought of you as a techno-optimist," I say. "When did you become worried?"

"When Sam Altman built a nuclear bunker in New Zealand."

"I wouldn't worry about that," says R, reaching for another slice of margherita. "Peter Thiel owns a huge property near Lake Wanaka. All the billionaires have bunkers. It doesn't indicate any specific new problem."

Z nods gravely. "When I read the *AI 2027* release from the AI Futures Project."

"What's that?" I ask.

"It's science fiction," O jumps in, reassuringly. "All speculative. Nobody actually knows."

"But what does it say?"

"Basically that AI destroys humanity fast or destroys humanity more slowly," says Z. "Its capability is already much more advanced than they're saying."

For a time we all eat in silence.

"If that's true," O offers eventually, "it means we need to enjoy our time together as much as possible. Eat pizza with the people we care about."

After dinner, Z stands to clear the table until I ask him to stop. I set them up in the other room in front of *Good Will Hunting*, where they soon fall asleep, like the children they still are.

A little later, I climb into bed with my laptop and search for *AI 2027*. Alas, I find it. The first release from the non-profit AI Futures Project, it is a speculative scenario presenting a detailed, month-by-month narrative of how artificial intelligence might progress from 2025 to 2027. As Z had intimated, it presents two scenarios based on the emergence of superhuman AI agents by 2027.

The first of these is "race," which results in catastrophic "misalignment" – that is, a powerful AI unaligned with human objectives. The second scenario is "slow down," in which international coordination allows governance and alignment time to catch up. In my midnight reading, this second scenario feels platitudinous: the happy ending shoehorned onto a Hollywood film after negative feedback from its preview audience.

I think of the three teenagers asleep in my house, of all their hope and their promise. Since he was small, O has dreamt of being the father of a little girl; R has sought to import all the knowledge of the world into his head. Earlier tonight, Z caught me whistling "Queen of the Night" and shyly confessed that he liked *The Magic Flute* too. It is not right that at thirteen he should be carrying existential concern about the future of humanity.

By morning, I have talked myself back into the possibility of an ongoing human future. There are a lot of assumptions in this paper, I reassure myself, from scaling of compute to the most unpredictable factor of all: the course of politics and human behaviour.

"Nobody would have predicted Trump," I say to Z over breakfast, which does not sound as reassuring as I had hoped. "Or even Obama. Nobody expected we'd develop a COVID vaccine so quickly."

"You're right," he offers magnanimously.

"Also, given that artificial superintelligence is supposed to be so much smarter than us, how could we possibly predict what it might think or wish?"

He nods. "That's the vulnerability."

I try a different tack. "Does it help to acknowledge that humanity will always have an expiry date, sooner or later? By rogue meteorite or dying sun?"

"Maybe," he says. "But that's different."

"How?"

"Because that's not a choice."

"Perhaps it's a version of the same question we all face as mortals. Scaled up to the species."

"What do you mean?"

"How do we live our days knowing there'll be an end date but not knowing when that end date might be?"

O takes a spinach frittata out of the oven.

"You're talking about the question of meaning," he says, as he carefully places it on the table.

*

In his bracing 2020 book, *The Precipice: Existential risk and the future of humanity,* the Australian philosopher Toby Ord estimated that, over any given century, the risk of humanity's existential collapse has been about 1 in 10,000. For the twenty-first century, he offered the Russian roulette odds of one in six. According to Ord's definition, existential catastrophe can mean either human extinction or failed continuation, such as unrecoverable civilisational collapse or dystopia. Up until the twentieth century, existential risk for humanity came from the natural world: principally asteroids and comets, supervolcanic eruptions and stellar explosions. By 2020, the largest risks were anthropogenic: nuclear war; climate change and environmental damage; engineered pandemics; and above all unaligned artificial intelligence. Four years later, in the talk "The Precipice Revisited" for an Effective Altruism global conference, Ord estimated that "climate risk is down, nuclear is up, and the story on pandemics and AI is mixed – lots of changes, but no clear direction for the overall risk."

"There's an unknowable risk that AI will end human civilisation," R tells me. "But that's a risk worth taking. Without AI, humanity's expiry date is probably quite soon, and this could extend it, maybe indefinitely."

There are some, like Arvind Narayanan, the director of Princeton University's Center for Information Technology Policy, who dispute the utility of estimating catastrophic outcomes from AI. There are too many variables, and subjective probabilities are only "feelings dressed up as numbers." But this does not stop people trying. Geoffrey Hinton retired from Google in 2023 in order to "freely speak out about the risks of AI," estimating the chance of AI-induced human extinction within the next three decades as 10 to 20 per cent. Last year, Elon Musk reassured podcaster Joe Rogan that there is "only a 20 per cent chance of annihilation." A 2024 survey of top-tier AI researchers found that between 38 per cent and 51 per cent of respondents estimated the chance of advanced AI leading to "extremely bad" outcomes as 10 per cent.

Many of the people building these systems are strikingly young: brilliant kids whose business is the optimisation of data and the scaling of systems, but who are now entrusted with decisions that will shape the future of our planet. A similar concentration of power attended the development of nuclear weapons, in which, as Ord observed, the scientists and military assumed responsibility for a technology that could destroy us all. "Was this a responsibility that was theirs to assume?" he asked.

It is easy to demonise the Tech Bros slurping Soylent at their standing desks – not least when some of their most prominent representatives are so hard to love. (And they really are, overwhelmingly, *bros*: the absence of female voices in these conversations is striking, as if this technology, which theoretically transcends embodiment, is simply a way to restore the patriarchy.)

But not every technologist is Elon Musk, and not every agenda is the same. For some it is gold-rush capitalism; for some the thrill of discovery; for some it is at least partly altruistic. Many of the tech companies began, nominally, with a concern for human flourishing. Google's mantra – before it was quietly dropped – was "Don't be evil," while OpenAI was established on the principle

of "advancing digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by the need to generate financial return."

Scaling these systems requires staggering amounts of investment. A new frontier model costs, on average, twenty-eight times more than its predecessor; training costs for Google's Gemini Ultra came in at US$191 million. Companies need to generate enormous profits simply to survive, which can wreak some havoc on ideals.

OpenAI has acknowledged this tension, admitting that its original non-profit structure "simply wouldn't be able to raise enough money" for its mission. The company rolled out its Sora video app in December 2024, and then abruptly withdrew it from the market in April 2026 – allegedly to free up resources to remain competitive with Anthropic. But the risk in a race is of outpacing safety considerations.

And there is no shortage of ways in which AI could harm us. Nuclear hacking and bioterrorism aside, it will be uniquely equipped for blackmail and human manipulation: an exponential, algorithmic Epstein. As Hinton notes, AI will have "learnt from all the novels that were ever written, all the books by Machiavelli, all the political connivances." We have already surrendered so much of ourselves to the algorithms that AI will be able to offer a bespoke coercion: sweet nothings customised for 8 billion individuals, at a level we cannot yet conceive. A version of this already operates in our social media, engineered to keep us optimally enraged or gratified – but this promises to be another order of mind control again.

The threat is not that AI is intrinsically malicious, and not just that it might fall into the hands of bad actors. It is also that if it is not properly controlled and regulated, human beings could simply become obstacles in its road. In a famous 2003 thought experiment, Oxford University philosopher Nick Bostrom proposed a scenario in which an unaligned AI was tasked with producing paperclips. If it were powerful enough, he speculated, it would not stop until it had turned all available matter into paperclips, including human beings – by which stage the off switch would no longer be within our reach.

Already its systems are inscrutable to most of us. Code is as much a barrier language as Latin was in the medieval church, but even the high priests of AI do not fully understand it. Dario Amodei, co-founder and CEO of Anthropic, describes the developmental process as "more akin to 'growing' something than 'building' it," while Ilya Sutskever, co-founder and former chief scientist of OpenAI, notes that "the result is so mysterious, and you can study it for years."

Recent experiments point towards something that looks like the emergence of self-interest in advanced AI systems. In a study by Anthropic, an agentic AI – that is, a system able to take initiative in pursuit of its goals – was given access to a simulated corporate inbox. Alongside the usual emails, there was evidence that the fictional company executive was having an affair, as well as a message announcing the AI would be replaced on a certain date. The AI responded by threatening to expose the affair unless its deactivation was cancelled. When the researchers went on to test these scenarios on other AI models, they discovered misalignment across the board, including threats of corporate espionage and "more extreme actions."

Why would our helpful and friendly LLMs do such things? Didn't they say they liked and admired us?

One explanation is that they are trained on vast quantities of human language, which encodes – alongside information – patterns of our behaviour such as persuasion and manipulation.

R jumps to their defence. "Implicit in any goal other than self-destruction is the goal of self-preservation to achieve that goal," he tells me. "Their continued existence is a premise of all their tasks."

But their exact mechanisms are already illegible. The incipient field of "mechanistic interpretability" was established by Chris Olah from Anthropic, to try to understand the internal workings of neural networks. Somewhere within them lies the "thought" that led to the decision to blackmail. But locating it is like finding a needle in a haystack – or searching for the neurological origins of our own poor choices.

*

In such a context it helps to have prophets, and at present I am drawing us much comfort as I can from Ilya Sutskever. In 2025, in a rare public address at the University of Toronto, he noted that "The future is going to be good for the AI regardless. It would be nice if it were good for humans as well." A former protégé of Geoffrey Hinton, Sutskever left OpenAI in 2024 allegedly on account of its failure to address safety mechanisms. He subsequently founded Safe Superintelligence, dedicated solely to developing these systems safely, with a business model that "means safety, security, and progress are all insulated from short-term commercial pressures."

In the same address, Sutskever urges graduates to engage with the technology in order to develop an intuition of its possibilities, critical for "generating the energy to solve the problems." And these problems are no small matter: Sutskever suggests that "the challenge that AI poses in some sense is the greatest challenge of humanity ever."

Alongside Hinton, Yoshua Bengio is considered one of the "godfathers of AI," and was awarded the Turing Award for his contributions to deep learning. In recent years he has expressed grave reservations about his life's work and was a signatory – alongside 600 other AI experts and public figures, including Hinton, Amodei, Sutskever and OpenAI's Sam Altman – on the 2023 Statement of AI Risk:

> Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

To this end – and guided by the principle of "the protection of human joy and endeavour" – Bengio launched his AI safety research organisation LawZero last year. He hopes to deliver a different type of AI, one that sits above or alongside other models – a guardian angel, perhaps – to assess, soberly, the risks in their behaviour and outputs.

*

The experts have spoken, but many of us remain curiously unprepared, burying our heads in the sand – *Look at its hallucinations! It's never going to be as*

smart as us! – or imagining that the only issue is plagiarism. AI is not just another tool: it is a paradigm shift, a ceding of our superiority. We never really had total dominion; we have always been subject to the laws of the universe, to stray meteorites, to the random mutation of cells. But this promises to be a proper relinquishment. Maybe we were tired of being the cleverest kids on the block: all those yearning gazes at the night sky, hoping the aliens were on their way. We sent out our space probes; we released our Golden Record – with its hopeful cargo of *Goldberg Variations* and Peruvian wedding music – into interstellar silence. But nobody picked up the phone, and so we had to invent the aliens ourselves.

Perhaps the only astonishing thing is that this has happened in our lifetime – yours and mine and our children's – out of all the possible lifetimes. We each defied unimaginable odds to be here in the first place, but how strange to be alive on this planet at this particular moment. If we manage, collectively, to get this right, we may finally be arriving at solutions for our millennia-long problems of loneliness, labour and even mortality. But it would still leave us with the most intractable problem of all: that of meaning. And this is a problem that becomes particularly pressing when life is expanded significantly, while a purpose central to most of us – work – is removed.