

# Power-hungry robots, space colonization, cyborgs: inside the bizarre world of 'longtermism'

J Oliver Conroy

Sun 20 Nov 2022 11.00 GMT



*Samuel Bankman-Fried, the founder  
and CEO of FTX, and an adopter of  
'longtermism'. Photograph: Saul  
Loeb/AFP/Getty Images*

Sam Bankman-Fried said his billions would save the world – but his philanthropic ideas ranged from the worthy to the severely outlandish

## **1. Most of us**

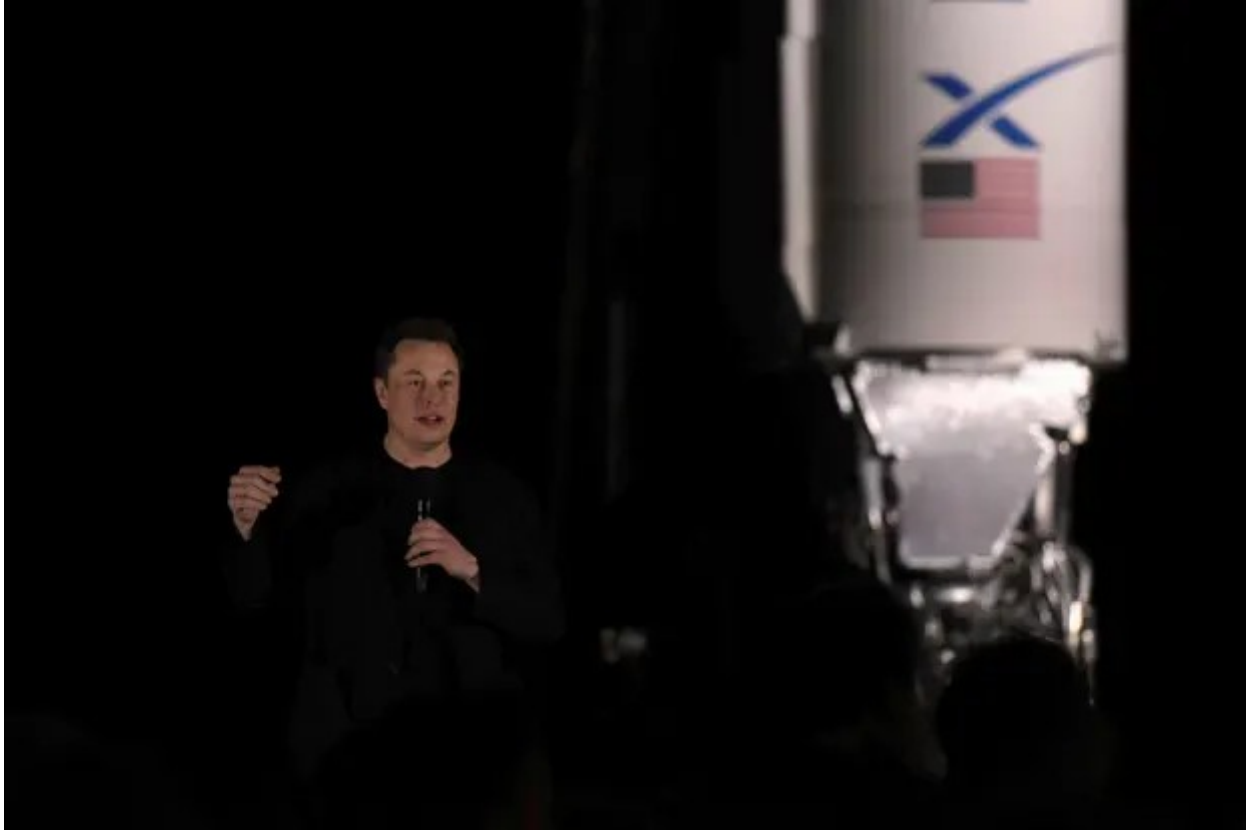
Most of us don't think of power-hungry killer robots as an imminent threat to humanity, especially when poverty and the climate crisis are already ravaging the Earth.

This wasn't the case for Sam Bankman-Fried and his followers, powerful actors who have embraced a school of thought within the effective altruism movement called “longtermism”.

In February, the Future Fund, a philanthropic organization endowed by the now-disgraced cryptocurrency entrepreneur, [announced](#) that it would be disbursing more than \$100m – and possibly up to \$1bn – this year on projects to “improve humanity’s long-term prospects”.

The slightly cryptic reference might have been a bit puzzling to those who think of philanthropy as funding homelessness charities and medical NGOs in the developing world. In fact, the Future Fund’s particular areas of interest include artificial intelligence, biological weapons and “space governance”, a mysterious term

referring to settling humans in space as a potential “watershed moment in human history”.



*SpaceX’s Elon Musk gives an update on the company’s Mars rocket Starship. Musk is a proponent of longtermism. Photograph: Callaghan O’Hare/Reuters*

Out-of-control artificial intelligence was another area of concern for Bankman-Fried – so much so that in September the Future Fund announced prizes of up to \$1.5m to anyone who could make a

persuasive estimate of the threat that unrestrained AI might pose to humanity. “We think artificial intelligence” is “the development most likely to dramatically alter the trajectory of humanity this century”, the Future Fund said. “With the help of advanced AI, we could make enormous progress toward ending global poverty, animal suffering, early death and debilitating disease.” But AI could also “acquire undesirable objectives and pursue power in unintended ways, causing humans to lose all or most of their influence over the future”.

Less than two months after the contest was announced, Bankman-Fried’s

\$32bn cryptocurrency [empire](#) had collapsed, much of the Future Fund's senior leadership had resigned and its AI prizes may never be rewarded.

Nor will most of the millions of dollars that Bankman-Fried had promised a constellation of charities and thinktanks affiliated with effective altruism, a once-obscure ethical movement that has become influential in [Silicon Valley](#) and the highest echelons of the international business and political worlds.

## **2. For Longtermists that the welfare of future humans is morally important**

Longtermists argue that the welfare of future humans is as morally important – or more important – than the lives of current ones, and that philanthropic resources should be allocated to predicting, and defending against, extinction-level threats to humanity.

But rather than giving out malaria nets or digging wells, longtermists prefer to allocate money to researching existential risk, or “x-risk”.

In his recent book *What We Owe the Future*, William MacAskill – a 35-year-

old moral philosopher at Oxford who has become the public intellectual face of effective altruism – makes a case for longtermism with a thought experiment about a hiker who accidentally shatters a glass bottle on a trail. A conscientious person, he holds, would immediately clean up the glass to avoid injuring the next hiker – whether that person comes in a week or in a century.

Similarly, MacAskill argues that the number of potential future humans, over many generations for the duration of the species, far outnumber the number currently alive; if we truly believe that all humans are equal, protecting future



humans is more important than protecting human lives today.

Some of longtermists' funding interests, such as nuclear nonproliferation and vaccine development, are fairly uncontroversial. Others are more outlandish: investing in space colonization, preventing the rise of power-hungry AI, cheating death through "life-extension" technology. A bundle of ideas known as "transhumanism" seeks to upgrade humanity by creating digital versions of humans, "bioengineering" human-machine cyborgs and the like.

People like the futurist Ray Kurzweil and his adherents believe that

biotechnology will soon “enable a union between humans and genuinely intelligent computers and AI systems”, Robin McKie [explained](#) in the Guardian in 2018. “The resulting human-machine mind will become free to roam a universe of its own creation, uploading itself at will onto a ‘suitably powerful computational substrate’,” and thereby creating a kind of immortality.

### **3. Feverish techno-utopianism distracts from pressing problems**

This feverish techno-utopianism distracts funders from pressing problems that already exist here on Earth, said Luke Kemp, a research associate at the University of Cambridge's Centre for the Study of Existential Risk who describes himself as an "EA-adjacent" critic of effective altruism. Left on the table, he says, are critical and credible threats that are happening right now, such as the climate crisis, natural pandemics and economic inequality.

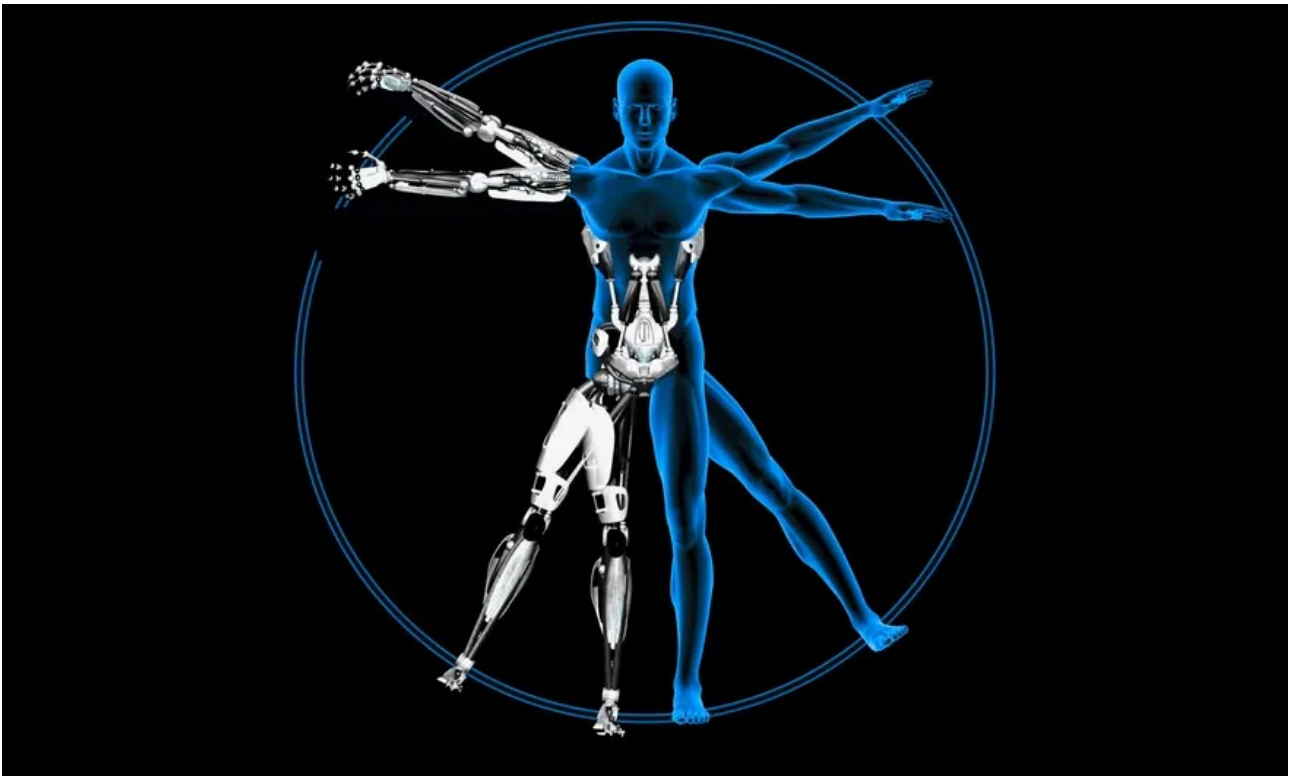
"The things they push tend to be things that Silicon Valley likes," Kemp said. They're the kinds of speculative,

futurist ideas that tech billionaires find intellectually exciting. “And they almost always focus on technological fixes” to human problems “rather than political or social ones”.

There are other objections. For one thing, lavishly expensive, experimental bioengineering would be accessible, especially initially, to “only a tiny sliver of humanity”, Kemp said; it could bring about a future caste system in which inequality is not only economic, but biological.

This thinking is also dangerously undemocratic, he argued. “These big decisions about the future of humanity should be decided by humanity. Not by

just a couple of white male philosophers at Oxford funded by billionaires. It is literally the most powerful, and least representative, strata of society imposing a particular vision of the future which suits them.”



*Some adherents of longtermism are interested in ‘transhumanism’, the idea that technology can extend our longevity. Composite: Lynsey Irvine/Getty*

Kemp added: “I don’t think EAs – or at least the EA leadership – care very much about democracy.” In its more dogmatic varieties, he said, longtermism is preoccupied with “rationality, hardcore utilitarianism, a pathological obsession with quantification and neoliberal economics”.

Organizations such as [80,000 Hours](#), a program for early-career professionals, tend to encourage would-be effective altruists into four main areas, Kemp said: AI research, research preparing for human-made pandemics, EA community-building and “global priorities research”, meaning the

question of how funding should be allocated.

The first two areas, though worthy of study, are “highly speculative”, Kemp said, and the second two are “self-serving”, since they channel money and energy back into the movement.

This year, the Future Fund [reports](#) having recommended grants to worthy-seeming projects as various as research on “the feasibility of inactivating viruses via electromagnetic radiation” (\$140,000); a project connecting children in India with online science, technology, engineering and mathematics education (\$200,000); research on “disease-neutralizing

therapeutic antibodies” (\$1.55m); and research on childhood lead exposure (\$400,000).

But much of the Future Fund’s largesse seems to have been invested in longtermism itself. It recommended \$1.2m to the Global Priorities Institute; \$3.9m to the Long Term Future Fund; \$2.9m to create a “longtermist coworking office in London”; \$3.9m to create a “longtermist coworking space in Berkeley”; \$700,000 to the Legal Priorities Project, a “longtermist legal research and field-building organization”; \$13.9m to the Centre for Effective Altruism; and \$15m to Longview [Philanthropy](#) to execute



“independent grantmaking on global priorities research, nuclear weapons policy, and other longtermist issues.”

Kemp argued that effective altruism and longtermism often seem to be working toward a kind of regulatory capture.



*Sam Bankman-Fried at a Senate agriculture, nutrition and forestry committee hearing in Washington DC. Photograph: Bloomberg/Getty Images*

“The long-term strategy is getting EAs and EA ideas into places like the

Pentagon, the White House, the British government and the UN” to influence public policy, he said.

There may be a silver lining in the timing of Bankman-Fried’s downfall.

“In a way, it’s good that it happened now rather than later,” Kemp said. “He was planning on spending huge amounts of money on elections. At one stage, he said he was planning to spend up to a billion dollars, which would have made him the biggest donor in US political history. Can you imagine if that amount of money contributed to a Democratic victory – and then turned out to have been based on fraud? In an already fragile and polarized society like the US? That would have been horrendous.”

## **4. The main tensions come from funding**

The main tension to the movement, as I see it, is one that many movements deal with,” said Benjamin Soskis, a historian of philanthropy and a senior research associate at the Urban Institute. “A movement that was primarily fueled by regular people – and their passions, and interests, and different kinds of provenance – attracted a number of very wealthy funders,” and came to be driven by “the funding decisions, and sometimes just the public identities, of people like SBF and Elon Musk and a few others”. (Soskis noted that he has received funding from Open Philanthropy, an EA-affiliated foundation.)

Effective altruism put Bankman-Fried, who lived in a luxury compound in the Bahamas, “on a pedestal, as this Corolla-driving, beanbag-sleeping, earning-to-give monk, which was clearly false”, Kemp said.

Soskis thinks that effective altruism has a natural appeal to people in tech and finance – who tend to have an analytical and calculating way of thinking about problems – and EA, like all movements, spreads through social and work networks.

Effective altruism is also attractive to wealthy people, Soskis believes, because it offers “a way to understand

the marginal value of additional dollars”, particularly when talking of “vast sums that can defy comprehension”. The movement’s focus on numbers (“[shut up and multiply](#)”) helps hyper-wealthy people understand more concretely what \$500m can do philanthropically versus, say, \$500,000 or \$50,000.

One positive outcome, he thinks, is that EA-influenced donors publicly discuss their philanthropic commitments and encourage others to make them.

Historically, Americans have tended to regard philanthropy as a private matter.

But there’s something “which I think you can’t escape”, Soskis said. Effective

altruism “isn’t premised on a strong critique of the way that money has been made. And elements of it were construed as understanding capitalism more generally as a positive force, and through a kind of consequentialist calculus. To some extent, it’s a safer landing spot for folks who want to sequester their philanthropic decisions from a broader political debate about the legitimacy of certain industries or ways of making money.”

Kemp said that it is rare to hear EAs, especially longtermists, discuss issues such as democracy and inequality.

“Honestly, I think that’s because it is something the donors don’t want us

talking about.” Cracking down on tax avoidance, for example, would lead to major donors “losing both power and wealth”.

The downfall of Bankman-Fried’s crypto empire, which has jeopardized the Future Fund and countless other longtermist organizations, may be revealing. Longtermists believe that future existential risks to humanity can be accurately calculated – yet, as the economist Tyler Cowen recently [pointed out](#), they couldn’t even predict the existential threat to their own flagship philanthropic organization.

There must be “soul-searching”, Soskis said. “Longtermism has a stain on it and

I'm not sure when or if it will be fully removed.”

“A billionaire is a billionaire,” the journalist Anand Giridharadas [wrote](#) recently on Twitter. His [2018 book](#) *Winners Take All* sharply criticized the idea that private philanthropy will solve human problems. “Stop believing in good billionaires. Start organizing toward a good society.”



Source of Guardian article:

<https://www.theguardian.com/technology/2022/nov/20/sam-bankman-fried-longtermism-effective-altruism-future-fund>

Wikipedia entry on FTX:

([https://en.wikipedia.org/wiki/FTX\\_\(company\)](https://en.wikipedia.org/wiki/FTX_(company)))

[Formatted for ebook-readers by  
Wergosum on 2022-11-21]