

# The do-gooder movement that shielded Sam Bankman-Fried from scrutiny

## Effective altruism, backed by Silicon Valley billionaires, now finds itself at a crossroads.

By [Nitasha Tiku](#)

The Washington Post

Updated November 17, 2022 at 12:25 p.m. EST|Published November 17, 2022 at 11:45 a.m. EST

When FTX founder Sam Bankman-Fried was a junior at MIT, he met William MacAskill, one of the founders of effective altruism, a philosophical movement that believes rationality is key to doing maximum good in the world.

[Tech is not your friend. We are. Sign up for The Tech Friend newsletter.](#)

For more than a decade, the EA community, as it's known, has attracted bright young students from elite colleges like Oxford and Stanford, encouraging them to take lucrative jobs in finance and tech so they can amass wealth and donate it to efficient charities. EA's emphasis on measurable impact has attracted billionaire tech philanthropists like Facebook co-founder [Dustin Moskovitz](#), who helped finance a tightknit web of nonprofits and academic institutes to grow the movement even as its priorities shifted from helping the world's poor to combating more theoretical risks, such as the rise of a hostile advanced artificial intelligence.

Bankman-Fried made an ideal recruit for MacAskill's "[earning-to-give](#)" pitch. Bankman-Fried had been raised as a utilitarian — a doctrine holding that the most ethical choice is the one that does the most good for the most people — and was already interested in protesting factory farming.

MacAskill, an Oxford philosophy professor, encouraged him instead to pursue a high-paying job in finance.

### [From the Magazine: The Rise of the Rational Do-Gooders](#)

As Bankman-Fried sought higher risks and rewards in cryptocurrency, launching the quantitative trading firm Alameda Research, the EA community continued to play a central role. The first people he hired for Alameda were from EA. The first \$50 million in funding came from an EA connection. And, for a period of time, half of Alameda's profits allegedly went to EA-related charities, according to [a profile of Bankman-Fried](#) commissioned by the venture capital firm Sequoia Capital, a significant investor in Bankman-Fried's cryptocurrency exchange FTX.

Over the past two weeks, a [bank run](#) on FTX exposed Bankman-Fried's alleged misuse of FTX customer funds to cover Alameda's debts, triggering a [bankruptcy filing](#), investigations by the U.S. Securities and Exchange Commission and the Department of Justice, and a cascade of chaos in the \$850 billion crypto market. It also vaporized Bankman-Fried's personal wealth, estimated at [\\$15.6 billion](#) as recently as Nov. 7, and is shining a spotlight on EA, an integral part of Bankman-Fried's origin story.

During Bankman-Fried's ascent, media portrayals invariably noted that the crypto wunderkind

drove a Toyota Corolla and planned to give his billions away, even as he courted celebrities and Washington power brokers. Indeed, his proximity to EA's brand of self-sacrificing overthinkers often helped deflect the kind of scrutiny that might otherwise greet an executive who got rich quick in an unregulated offshore industry.

### [Sam Bankman-Fried charmed Washington. Then his crypto empire imploded.](#)

Now EA is at a crossroads. Money expected to fund the next phase of growth has evaporated, while questions have arisen about whether money already donated to speculative EA projects was unethically obtained. EA leaders also [face questions](#) about what they knew about the business dealings of a billionaire whose reputation they helped burnish. Meanwhile FTX's collapse has raised existential concerns: In its current state, would EA survive its own calculation as a force for good?

On Nov. 11, the day FTX filed for bankruptcy, MacAskill said in a Twitter thread: "For years, the EA community has emphasized the importance of integrity, honesty, and the respect of common-sense moral constraints. If customer funds were misused, then Sam did not listen; he must have thought he was above such considerations. A clear-thinking EA should strongly oppose 'ends justify the means' reasoning."

"If this is what happened," MacAskill continued, "then I cannot in words convey how strongly I condemn what they did. I had put my trust in Sam, and if he lied and misused customer funds he betrayed me, just as he betrayed his customers, his employees, his investors, & the communities he was a part of."

MacAskill nor Bankman-Fried responded to requests for comment.

Philosopher Émile P. Torres, one of the movement's [harshest critics](#), suggested that the FTX implosion "might trigger some serious reorganizing of the community." But, they added: "It's hard to imagine EA bouncing back from this easily."

Born at Oxford, EA is a community of roughly 7,000 adherents — largely young, White men connected to elite schools in the United States and Britain, according to recent [annual EA surveys](#). Prominent on college campuses, the ideology also has taken hold in fields like artificial intelligence, where it has reshaped industry norms. Before Bankman-Fried's empire unraveled, EA had access to an estimated \$46 billion in funding and was making a strategic push to influence global public policy.

On a recent podcast about the movement's inroads [at the United Nations](#), MacAskill said he hoped to make his ideas for humanity's priorities "something that people in positions of power can take seriously."

EA adherents are "in journalism, they're in academia, they're in Big Tech, and they're coordinating around this idea of being value-aligned," said Carla Zoe Cremer, a PhD student at Oxford and former research scholar with the Future of Humanity Institute. Those power centers are ideal "if you just want to get s--- done," Cremer said. "The question is, what do they want to get done?"

An EA critic, Cremer says the movement has yet to figure that out. Instead, she said, it is taking the more dangerous approach of amassing power and then figuring out what to do with it.

## **From bed nets to Skynet**

The name effective altruism was coined in 2011 as an umbrella term for disparate efforts, like the

charity GiveWell, to more rigorously evaluate international aid and to encourage effective giving through nonprofits like Giving What We Can and 80,000 Hours.

Its underlying philosophy marries 18th-century utilitarianism with the more modern argument that people in rich nations should donate disposable income to help the global poor. But there's also a big emphasis on math, borrowing from economics and decision theory to prioritize causes and measure potential improvements in quality of life. Early on, this cost-benefit approach produced a lot of donations for mosquito nets to prevent malaria in susceptible countries.

From the beginning, however, the do-gooder group in Oxford was tied to a similar subculture in the Bay Area. The leading intellectuals of that world, like AI theorist Eliezer Yudkowsky, wrote for an online forum called Less Wrong, which, like effective altruism, also attracted a community of young people more interested in specific modes of argumentation than politics.

Yudkowsky and Nick Bostrom, who was also at Oxford, shared a similar theoretical concern about AI development: Specifically, that once artificial intelligence became as smart as humans, things could quickly spin out of control. Their ideas might have remained a mostly intellectual exercise, worked out in white papers and online forums, if not for a handful of Silicon Valley tycoons who elevated them to a bigger stage.

[Cari Tuna and Dustin Moskowitz: Young Silicon Valley billionaires pioneer new approach to philanthropy](#)

Open Philanthropy, the primary philanthropic funding vehicle for Moskowitz and Holden Karnofsky, a former hedge fund trader who helped kick-start EA, had long backed causes in global health and development. But things began to heat up around 2015 when Elon Musk donated \$10 million to the Future of Life Institute to fund an AI safety research program. Bill Gates called the program “absolutely fantastic” and said he shared Musk’s concerns about AI.

Musk has called AI an “existential risk,” citing “Superintelligence,” the best-selling book by Bostrom. But there’s an asterisk on “existential”: Bostrom lays out a long-term vision for a techno-utopia, millions or billions of years in the future, where we colonize space and harness the power of the stars to upload our consciousness, evolving into some kind of “digital people.” In Bostrom’s view, “existential risk” is anything that stands in the way of this utopia, meaning he sees the nonexistence of computer-simulated people as a moral tragedy. In the extremist view, it’s on equal footing with the death of someone alive today.

Both funders and philosophers arrived a similar conclusion: AI’s evolution is inevitable, and making sure it remains friendly to humanity is a top priority.

[Elon Musk: ‘With artificial intelligence we are summoning the demon.’](#)

With the imprimatur of tech billionaires, there was something of a Cambrian explosion of EA organizations: the Global Priorities Institute, the Forethought Foundation, EA Funds, The Longtermism Fund, Longview Philanthropy and a revolving door between many of them, with nonprofit directors moving to granting organizations and grant advisers serving as board members on organizations receiving funds.

Open Philanthropy has donated the most money, \$728.8 million, to global health and development. But it also has donated \$234 million to Effective Altruism Community Growth and \$255.8 million

to fight potential risks from advanced AI. That's compared with \$4.9 million on, for example, South Asian air quality.

## **Altruism, Inc.**

On elite campuses, students might receive a free copy of books like MacAskill's "Doing Good Better" or Toby Ord's "The Precipice." They might be invited to lectures, to study at the university's EA co-working space, to set up free career counseling with 80,000 Hours, to attend a coaching start-up co-founded by MacAskill, or to get funding to pursue "longtermist" research with EA Grants. There are EA workplace groups for employees at Microsoft, Palantir and Amazon and even an EA group devoted to writing Wikipedia articles about EA.

When the movement's focus changed, Cremer said, the community put increasing emphasis on what it calls being "value-aligned," an ill-defined concept that more and more has been used to define in-group status. There is a shared set of source texts, shared style of speaking and shared mannerisms. Cremer says it facilitates deep trust between EA members, which can give rise to such habits as prizing EA-alignment over technical expertise and tolerating certain conflicts of interest.

### [House to probe FTX collapse as Biden administration warns of crypto risks](#)

Bankman-Fried had pledged to donate his billions to EA causes — specifically to the existential risks that have become the movement's focus and that are outlined in MacAskill's new book, "What We Owe the Future." In February, Bankman-Fried announced the FTX Future Fund, naming MacAskill as an adviser.

EA institutions bolstered Bankman-Fried's image as a self-sacrificing ethicist. In an interview on the [80,000 Hours podcast](#), host Rob Wiblin laughingly dismissed the idea that Bankman-Fried's pledge to donate his wealth was insincere.

"Are there any fancy expensive things that you are tempted by on a selfish level?" Wiblin asked. "Or is it just nothing?"

"I don't know, I kind of like nice apartments," said Bankman-Fried, who lived until recently in an oceanfront penthouse in the Bahamas.

Among the beneficiaries of Bankman-Fried's philanthropy were a clutch of Democratic congressional candidates; he was a top party donor in the midterms, spending nearly \$36 million, according to [Open Secrets](#). The bulk of it, [\\$27 million](#), went through Protect Our Future Pac, which supported candidates who prioritize preventing pandemics, a major focus of longtermers.

### [Before FTX collapse, founder poured millions into pandemic prevention](#)

The influx of funds into EA mirrored the mood around Silicon Valley start-ups where, until recently, easy money chased too few good ideas, but no one really wanted the party to end. Nick Beckstead, CEO of the FTX Future Fund, chastised community members for thinking they would get a blank check. "Some people seem to think that our procedure for approving grants is roughly 'YO-LO #sendit,'" Beckstead wrote on the Effective Altruism forum in May. In December 2021, another forum poster wondered whether EA was suffering from "TMM," short for "Too Much Money."

Many of the Future Fund's grants went to growing the movement. The fund shoveled millions of dollars to EA and longtermist co-working spaces and millions more to a fellowship for high-schoolers. But the largest amounts went to the Centre for Effective Altruism and Longview

Philanthropy, where MacAskill and Beckstead are advisers.

Then the money stopped. FTX's sudden demise came two years after [legal troubles](#) befell another crypto billionaire who had pledged funds to the cause: BitMEX's Ben Delo, who was [sentenced](#) in June to 30 months' probation for flouting anti-money-laundering statutes.

### [FTX investors sue Sam Bankman-Fried and celebrity endorsers](#)

Bankman-Fried, whose face was plastered across city billboards and whose logo was on major sports arenas, won't be as easy to erase. But he already appears to be taking steps toward protecting EA. On Tuesday, he denied he was ever truly an adherent and suggested that his much-discussed ethical persona was essentially a scam.

"I had to be [ethical]. It's what reputations are made of, to some extent," [he told Vox via Twitter direct message](#). "I feel bad for those who get f---ed by it, by this dumb game we woke Westerners play where we say all the right shiboleths [sic] and so everyone likes us."