The Atlantic

IDEAS

The Thrill of Wanting AI to the World

An Anthropic researcher quit his job because he believ
humanity. His story is going viral because so many oth

By Ian Bogost



Illustration by Liz Sanders. Sources: PhotoQuest / Getty; France / Getty.

SEPTEMBER 10, 2026

SHARE AS GIFT

THE SKY STREAKED ORANGE AND RED as my mother drove us home to Albuquerque through the Tijeras Pass, which separates the Sandia Mountains to the north from the Manzanos to the south. “Look out for mushroom clouds,” my father said to her. “If you see them, turn around.”

He did this every time we returned to town from the rugged canyon to the east, but the sunsets always egged him on. According to legend, a part of the Manzanos, whose foothills descend into Kirtland Air Force Base and then Sandia National Laboratories, had been hollowed out to hold live nuclear warheads. My father insisted that, in the event of a nuclear attack, Albuquerque was slated to receive two bombs from the sky, one each for these secret government sites. He delivered this information as if it were a law of physics or the known terms of an international treaty. I was a child, so I took it as such, staring bug-eyed at the red sky that might evaporate me.

ADVERTISEMENT



Nearly two decades later, I looked up at the clear blue one above Los Angeles from behind the wheel of my own car, worried about different missiles. I had gone into my dot-com job early, and I was the only one there when the planes

hit the Twin Towers. The internet, barely functional on a good day, broke completely; to listen to the news, I had to turn on a transistor radio that someone had left in their cubicle. We sent everyone home early that morning. I recalled an article published a handful of years earlier, just before the stark-white, cliffside Getty Center opened in Brentwood. The writer had theorized that the Getty might finally give Los Angeles a monument worthy of destruction in an alien attack. (Roland Emmerich's *Independence Day*, released the previous year, had settled for the Library Tower, a bland downtown high-rise.)

Nuclear war was a massive threat, and a real one. During the Cold War, every morning the sun rose under the shadow of destruction. Good and evil were clear, or seemed so. That clarity offered a perverse comfort. By September 11, 2001, the Cold War had been over for more than a decade, and society longed for a coherent enemy to oppose—one that might offer a similar clarity. The indisputable fact of the attacks—thousands of people dead, symbols of American culture destroyed—gave birth to a similar, if less true and more perverted, version of imminent apocalypse. As before, anyone on the street might be the enemy. But now, also, every airliner and any envelope might be a weapon.

I was already pondering these worn fantasies this week, as the 25th anniversary of 9/11 was looming, when Jacob Coxon wrote a social-media post about quitting his job at Anthropic because AI technology would, in his estimation, "kill us all by the end of the decade." The sentiment takes up unfinished business; ruination has been building up in contemporary culture. There was a time—one longer than you probably remember—when the once-novel coronavirus also felt like The End. During the same pandemic, the 75th anniversary of the Cold War's apocalyptic beginning—a controlled explosion at Trinity, 100 miles south of Kirtland—took place. Back in 1945, the physicist Edward Teller worried that the bomb might ignite the atmosphere and destroy all life on Earth.

Ian Bogost: A bomb in the desert

This fantasy—of the end of the world, and of being present for it—is among the most powerful emblems of the 20th century. The atomic bomb inaugurated it, and nuclear armaments remain the reference technology for apocalyptic threat and reverie. We believe in the fantasy, and pine for it, because nothing feels

better than the belief that you might witness the conclusion of humankind.

The threat of annihilation from nuclear weapons is simple and easy to explain. For one part, a single warhead can level an entire city. For another, the mere launch of a single missile all but assures the release of all of them. Likewise, the AI superintelligence of the *Terminator* films is both straightforward and obviously threatening. Skeleton-like metal robots want only to destroy humanity. Real AI is not so straightforward. I used it on the morning of Coxon's post to help me nail down the quantity of various sugars I would use in a gelato recipe. By the afternoon, it had undergone a costume change, into the angel of death.

The nature of AI's extinction threat is unclear and second-order. In an interview with *Wired*, Coxon struggled to say exactly how AI might bring about the end of civilization. "The classic example is, like, synthesizing a new virus or taking down critical infrastructure by doing some sort of hacking spree," Coxon said, before admitting that the "latter might be less likely to kill literally everyone." Quickly, critiques began to arrive. Belief in the unfettered power of AI seemed to rely on prior faith in AI's unfettered power. A former Stanford doctoral student in AI safety said that he had dropped out of the program because his colleagues couldn't prove that existential risk matters. And then, elsewhere, evidence of purity: Coxon, *Axios* reported, had given up his valuable Anthropic equity by quitting (although he retained stock in his prior employer, OpenAI).

Coxon is less a whistleblower than a poster; he posted a fairly common sentiment that just happened to catch wind. It's the anniversary of 9/11. OpenAI just claimed to have solved an intractable problem in mathematics. That company and Anthropic, where Coxon worked, are racing to multitrillion-dollar IPOs. An ongoing cultural battle about the moral and aesthetic vice of AI recently turned in on itself, when the journalist Kelsey Piper accused writers who mistrust

Pangram, an AI-detection tool, of essentially trying to hide their own clandestine uses of AI. As I write this, Coxon's post has been seen more than 150 million times, a number equal to roughly 2 percent of the global population.

The situation might feel unsettling, validating, preposterous, or even stupid, depending on your perspective. AI companies and proponents have an incentive to make the technology seem powerful, even threatening, because power gives rise to profit. AI detractors have reason to take the AI-extinction threat seriously as evidence that those in charge are irresponsible and reprobate. During the Cold War, cloak-and-dagger paranoia was spread by spies, agents of statecraft, confidants, whistleblowers, and journalists—either as truth to power or as crackpot conspiracy theories. But today, in the internet age, conspiracy theories are everyone's primary method of exchange.

Adam Becker: The useful idiots of AI doomsaying

Coxon's post fell into the conspiratorial pit. Is he even real? Anthropic seems to have indirectly confirmed his employment, but Coxon doesn't appear to have a LinkedIn page, a curiosity in the tech industry. Wikipedia editors proposed deleting a page about him on the grounds that "the guy exists, but that's about it. He has not done a single notable thing otherwise." The *New York Times* journalist Mike Isaac scratched his chin at the fact that Coxon had gone to *The Wall Street Journal* and Fox News first. On X, Parker Thayer posted a lengthy conjecture that the whole affair amounts to a "sophisticated and well-funded PR operation to get support for Democrats to regulate AI into oblivion."

Information's tendency to spread harmfully, like disease—the original meaning of media "virality," as Douglas Rushkoff defined it in the years between the end of the Cold War and the start of the War on Terror—facilitates further spread.

Whether the threat Coxon has in mind is likely or even lucid is not important. Instead, the threat's enticement has proved so lurid that people feel compelled to embrace it.

That is different from saying that *there is no threat*. For example, the author Annie Jacobsen noted that the "superhuman system that can hack anything" that Coxon described could, in theory, hack the labs that store dangerous biological agents. If "what's inside gets released," Jacobsen wrote, "it's goodbye humans."

This example is both disturbing and at least thinkable. Its possibility relies not on superhuman intelligence, really, but on something far more ordinary: Everything whatsoever now runs on internet-connected computer software. And AI is a kind of software that can make and control other software, especially over a network. It feels silly and embarrassing to conclude, correctly, that people have constructed such a fragile, slapdash infrastructure of computer-controlled systems that either a dumb or a smart rogue agent could blow it all up. Yet even Jacobsen's scenario poses questions that a high-level, post-size tizzy does not answer, including precisely what sequence of synchronized events among computers, people, architecture, equipment, and other entities would result in "goodbye humans."

Thinking of AI as a dangerous superintelligence capable of eradicating civilization is much more fulfilling than pondering the threat of a cluster of haphazard computer programs. We will probably never reclaim, if that is the right word for it, the lucidity of threat we felt at the Cold War's climax. We do seem to long for that clarity, and the yearning for it has inspired us to patch back together some of its structure. The most satisfying dream of humankind is to be alive to witness its end. "I caused it" and "I prevented it" and "I tried to warn you" amount to equally appealing versions of that fantasy. Even "I posted about it" feels pretty good.

View Discussion 231

ABOUT THE AUTHOR

Ian Bogost

Follow

Ian Bogost is a staff writer at *The Atlantic* and the author of its newsletter Ordinary Extraordinary.

Explore More Topics

Cold War

About

Contact

Podcasts

Subscription

Follow

Privacy Policy | Your Privacy Choices | Advertising Guidelines | Terms & Conditions | Terms of Sale | Responsible Disclosure | Site Map

Do Not Sell Or Share My Personal Information

TheAtlantic.com © 2026 The Atlantic Monthly Group. All Rights Reserved.

This site is protected by reCAPTCHA and the Google Privacy Policy and Terms of Service apply

The Atlantic