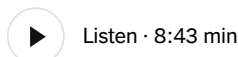


GUEST ESSAY

This Is How We Get Moral A.I. Companies

April 26, 2026

**By Paul Ford**

Mr. Ford is an essayist and a technologist.

See more of our coverage in your search results.

[Add The New York Times on Google ↗](#)

Artificial intelligence can be wondrous, but the technology underneath is more than a little monstrous. It eats up all the words in the world, from blogs to books, often without permission. It burns whole forests' worth of energy, digesting that raw material into its models, and gulps billions of gallons of water to cool down. These are the same qualities we perceive in Godzilla, but distributed. Is it any wonder that the Japanese word “kaiju,” or strange beast, has “AI” smack in the middle?

Mere greed didn't get us here. In fact, ethics did. The big A.I. labs' starry-eyed founders believed that the only way to stop the looming threat of a superintelligence that might kill us all was to create an aligned A.I. that would remain fond of humans. A friendly Godzilla could stop bad Godzillas before they got to Tokyo Bay. Sam Altman, Elon Musk and others came together to build the world's defense squad, which they called OpenAI. They built safety teams on which employees spent their days poking at the Godzilla eggs (testing chatbots) to make sure they wouldn't kill everyone when hatched. (One of the heads of those OpenAI safety teams was Dario Amodei, who left in 2020 to help found an even more aligned company, Anthropic.)

Companies are companies. They will, eventually, be expected to turn a profit. Humanistic goals will become subsumed by data-driven metrics. The idea of doing good brings everyone together — but somehow, “good” ends up a conflicted border, with angry people on either side. An A.I. company may want to do good, but it cannot do so on its own. It needs to be guided through rules and regulations.

A good portion of the earlier crop of A.I. thinking came out of the effective altruism movement, which calls for maximizing the good you can do by pursuing research-driven philanthropy. (One prominent practitioner is the incarcerated crypto entrepreneur Sam Bankman-Fried.) It's not a simple credo, but a big part of the ideology is oriented around preventing long-term threats to

humanity. For instance, if you believe that A.I. can become self-aware enough to modify itself and get smarter, then you might imagine it would modify itself into a state of total control of all the world's digital resources.

What can you, a bright, nerdy person, do to stop this? Only one thing: Build your superintelligence first and make it good, like you. Whatever your methods, they will seem valid — spinning up crypto schemes, possibly breaking copyright laws so you can feed your model every pirated text on the internet, blowing a hole in the labor market or raising the earth's temperature. Preventing an A.I. calamity is just that important.

Sign up for the Opinion Today newsletter Get expert analysis of the news and a guide to the big ideas shaping the world every weekday morning. [Get it sent to your inbox.](#)

Ten years ago, this was normal Silicon Valley conversation — solipsistic and nerdy, a big, expensive thought experiment in which people like Mr. Musk and Stephen Hawking would opine about how A.I. must be built to serve humans.

Then in late 2022, people started to realize that ChatGPT could do their homework. The money exploded.

Fast-forward to the present: Nearly a billion people are using ChatGPT. What appears to be a multitrillion-dollar economic bomb has been unleashed into the world. The loudest voices believe A.I. will either demolish the labor market — creating a jobless dystopia or creating a (similarly jobless) utopia — or reveal itself soon as a huge fraud and a bubble.

With every passing week, there are new fights over what these companies should and should not be allowed to do. Anthropic is suing much of the federal government, including the Defense Department; Mr. Musk is suing OpenAI; The New York Times is suing OpenAI and Microsoft. The nest of A.I.-related lawsuits is so vast that you need A.I. to keep track of it. Social outcry is at a fever pitch, often focused on the rise of data centers increasingly peppering the American map. Someone firebombed Mr. Altman's family home. This is, without irony, a disturbing outcome for an ethical movement designed to protect humanity.

Over three decades of watching the tech industry and watching big companies grow from tiny teams to global powers, I've observed the same pattern: Ethics don't scale up. Tech companies like to start with a mission. Google wanted to connect the world's information; Microsoft wanted to put a computer on every desktop; Twitter wanted to give all people a platform to publish their thoughts. These are good ideas — the stuff of TED Talks. But users show up with their own beliefs and ideas, by the millions. As a tech founder, you end up putting enormous work into making users behave (and stopping them from breaking the law). Lawsuits pour in, saying you did wrong, some because you're a convenient target.

All the while, money keeps gushing in. You start out transparent, sharing your journey, but then before an initial public offering of shares, you must honor the S.E.C.-mandated quiet period and restrict promotional communications. After that, the transparency never quite returns. The market demands a rising stock price. Your company still makes a lot of software, but a huge

amount of time goes to tax strategy and compliance.

At that scale, people start to blur together, and human users can become aggregate pools of statistics and growth vectors that go up and down — a mulch into which you plant your products.

The entire culture of American technology is built around two terms: disruption and, of course, scale. But ethics are constraints on disruption and scale. Truly ethics-bound organizations — the U.S. justice system, the American Medical Association, the Catholic priesthood — have hard scaling limits. Their rules run deep, and their requirements to serve are so onerous that only a few people can do the job. Punishments for transgressors include losing their licenses, being defrocked and being disbarred. Software industry people might have good degrees and are often good people, but they are making it up as they go along. They take no oath, are inconsistently certified and can only be fired, not exiled from the trade.

OpenAI set out to be inherently good — a dot-org. But it stumbled into a seam of pure digital gold in the form of large language models. To develop that technology further, it has made a painful, awkward transition to being a dot-com. (OpenAI says the for-profit arm continues to be overseen by the original nonprofit entity.) The subsequent level of drama has been difficult to behold. A few years ago, Mr. Altman publicly called for industry regulation, and he still does, but OpenAI has also lobbied against it — for example, supporting an Illinois bill that, if it becomes law, will limit the liability of A.I. companies in mass deaths.

But regulation is absolutely in the interests of both America and the big A.I. companies themselves. Let me add two more terms people should know: “Google zero” and “model collapse.” Google zero (coined by Nilay Patel, the editor in chief of The Verge) is when Google stops sending traffic to websites and just provides an A.I. answer instead. When that happens, websites get less traffic, sell fewer ads and make less money. As a result, they may not be able to produce as much content. Model collapse is related: It’s when the A.I. models run out of knowledge to digest. What then? Do they excrete their own prose to redigest? Do they just give up?

Silicon Valley types like to say that data is the new oil. I think that’s right in two ways: Data is valuable, but it’s also a commodity, and these new A.I. tools are infrastructure. We regulate the electric grid, so why not these?

In this new world, there are so many new things to regulate: Deepfakes, A.I. liability, copyright rules, model bias concerns and ecological costs top the list. And we will also need to protect the digital commons and incentivize people to write and do things online. So there will need to be a very long A.I. bill, and Congress will probably use ChatGPT to write it. At the risk of overstepping, I’d call it the Keeping American Ingenuity, Jobs and Unity Act. Or, the KAIJU Act. I hope that eventual bill’s authors consider restoring the web a little, like a wetlands — if for no other reason than we should be feeding our Godzillas healthy discourse, to help settle their nuclear heartburn.

Paul Ford is an essayist and a technologist. He is a founder and the president of Aboard, an A.I.-powered software acceleration platform.

The Times is committed to publishing a diversity of letters to the editor. We’d like to hear what you think about this or any of our articles. Here are some tips. And here’s our email: letters@nytimes.com.

Follow the New York Times Opinion section on Facebook, Instagram, TikTok, Bluesky, WhatsApp and Threads.

A version of this article appears in print on , Section SR, Page 8 of the New York edition with the headline: Is There Such a Thing as a Good A.I. Company?