The New York Times https://www.nytimes.com/2026/06/17/opinion/ai-dangerous-openai-anthropic.html

GUEST ESSAY

# Dear A.I. Companies, the Doom Trolling Needs to Stop

June 17, 2026

Listen · 10:42 min

**By Cal Newport**
Mr. Newport is a professor of computer science at Georgetown University and the author of "Deep Work."



Technology revolutions in the digital age are typically accompanied by optimism and excitement — recall Steve Jobs basking in thunderous applause as he introduced the iPhone in 2007. The major A.I. companies seem to be following a darker and weirder strategy: They like to solemnly describe the harms that their models will cause, while acting helpless to do anything about it.

Anthropic recently dropped a classic of the form: a scary-sounding report titled "When A.I. Builds Itself" that claims A.I. could be moving closer to the capability of "autonomously designing and developing its own successor." The company hopes that this recursive self-improvement will bring "enormous good" to the world, but also openly worries it might lead to humans "losing control" of these systems.

The public reaction to this report focused on a section that seemed to call for a worldwide pause on A.I. development. But if you read more carefully, it becomes clear that a pause isn't actually what Anthropic proposes. Its report says that "if it were possible" to slow down the technology, then we should, but so long as the "least cautious" actors were advancing full speed, it suggests that Anthropic will have no choice but to do the same.

Like a cat leaving a dead bird at your doorstep, Anthropic catalogs the grim future that its products might produce, shrugs its shoulders and then returns to its furious efforts

to make these warnings a reality.

Anthropic isn't alone in this nihilism. Sam Altman, the chief executive of OpenAI, frequently makes bleak claims. He compared the company's large-language-model efforts to the development of the atomic bomb, and he posted an image on social media of the Death Star from "Star Wars" in conjunction with the release of GPT-5. He's also argued, many times, that the only response to the economic damage that A.I. tools will inevitably produce would be to institute wide-scale financial support from the government, such as a universal basic income or a public wealth fund.

Let's call this strategy "doom trolling." It's one of the defining and most arresting properties of our current A.I. moment, and I've come to believe that it's morally indefensible.



**There are really only two options for the intentions of A.I. companies when they engage in doom trolling**. The first is that they actually believe that the systems they're building have a nontrivial chance of producing hugely disruptive events — from destroying the economy in the best case to wiping out our species in the worst. If this were true, every reasonable ethical system would argue that there is only one acceptable response: to immediately stop working on any product that might accelerate such a future, and lobby with all of your resources to help force other A.I. companies to do the same. From a moral perspective, any other reaction would be monstrous.

The second option is that these A.I. companies aren't really concerned about these risks, and that they're injecting these doses of unresolvable doom for other reasons. They might want to amplify the perceived power of their technology at a time when they're setting up their initial public offerings. Or they hope their performative reports and somber interviews will help them compete for top engineering talent coming from a Silicon Valley culture that's steeped in this type of doomerism. The venture capitalist and A.I. adviser David Sacks recently suggested that Anthropic was using fear-mongering tactics as a method of "regulatory capture," which can impede upstart competitors. Any of these reasons would mean that these companies are laundering the anxiety of millions to improve the financial fortunes of a vanishingly small number of major stockholders. This cynicism would be equally monstrous.

When it comes to A.I., we've become so used to this tone of helpless, stress-inducing

prognostication that we've lost sight of its strangeness. Imagine if Ford put out a report saying that it feared its popular F-150 trucks might soon start bursting into flames, but that there was nothing the company could do about it because automotive technology was too inevitable and important to slow down. You're probably struggling to picture this scenario because no reasonable consumer product company would ever act like this.

The A.I. companies could start behaving the same way. To do so would require that they stop treating A.I. like some inevitable force that they're struggling to steward. It's not. It's a collection of specific tools that these companies are choosing to design and sell according to specific business plans. Accordingly, they need to talk about their offerings like any other consumer product. This means explaining clearly whom these products are for, justifying their benefits and, critically, taking full responsibility for any harm they might cause. Just because A.I. currently enjoys a high-tech sheen doesn't make it exceptional with respect to common-sense safety standards.

If these A.I. companies insist on continuing to pretend that they're merely stoic observers of an unavoidable dystopian future, then perhaps it's time to force the issue. As consumers, we can refuse to play the doom-trolling game. Next time Anthropic releases a dire report, or Sam Altman's voice cracks as he imagines the disruption that OpenAI is unleashing, we can pivot back to the pragmatic: "OK, but what benefits am I getting by spending $1,000 a month on tokens?" If they continue to ratchet up the doom, then perhaps it's time to transform dread into ridicule: The earnest pseudoscience of Anthropic's white papers already borders on satire. The aura surrounding A.I. encourages a fretful submission to these tech leaders, but this could rapidly change.

(The New York Times has sued OpenAI and its partner, Microsoft, accusing them of copyright infringement of news content related to A.I. systems. OpenAI and Microsoft have denied those claims.)

The government can play a role here as well. President Trump recently signed an executive order that enabled A.I. companies to submit, on a voluntary basis, their models for risk assessment before their release. This program should be made mandatory, and the government should be willing to call the bluff of any company bragging about the destructive potential of its products, ending the era in which these A.I. labs can simultaneously terrify the public about their technology while continuing to develop and market it without constraint.

Remarkably, the current administration, in its own capricious and inscrutable way, might be lurching in this direction. In April, Anthropic announced that it would not publicly release its new Claude Mythos Preview L.L.M. because its ability to find and exploit software bugs could create "severe" fallout for our economy and safety. It

decided to share it with only a few organizations so that they can patch security vulnerabilities — a campaign that prompted great fear and consternation, but that also bolstered Anthropic's brand as a safety-conscious leader in A.I. technology.

Last week, the company released a version of the model now protected with "guardrails" for general use and another with lifted safeguards in some areas that it provided to a small group. The Trump administration, citing national security concerns, surprised the industry by pushing back. It put the models on an export-control list, which forced Anthropic to temporarily disable access. The jaded interpretation of Anthropic's actions is that its previous hand-wringing about Mythos was a publicity stunt (it is not clear why the "guardrails" they added last week could not have been added back in April when the model was first announced), which casts the administration's actions, in some part, as calling foul on Anthropic.

The frontier labs like to cite China to justify moving as quickly as possible on their products, free from any meaningful regulatory pressure (though some, like Anthropic's chief executive, Dario Amodei, have called for regulations). But managing geopolitical risks is the job of the government, not Silicon Valley. When the American biotech industry became worried about the negative potential of genetic engineering tools in the 1990s, it didn't rush ahead to clone humans before China could get there; it instead lobbied Congress and relevant international bodies to limit the most odious possibilities for these innovations. The possibility that another actor might do something bad doesn't give you the moral right to do something similar.

The courts can also become a source of pressure toward changing how A.I. companies talk about their products. A decade ago, the social media titans found themselves in a similar position to our current A.I. leaders. They attempted to abdicate responsibility for the obvious negative impacts created by their apps by treating them as something too fundamental to be restricted: a digital town square that represented the natural evolution of communication and democracy.

But in a landmark verdict in March, a jury found Meta and Google liable for millions of dollars in damages for harms caused by their social media platforms. Many hundreds of similar lawsuits are now making their way through the court system, representing a major threat to these companies that spent years profiting with impunity from their addictive, brain-warping designs. Could A.I. face similar litigation? We got the first hint that they might last week when a court in Germany ruled that L.L.M. operators are liable for the text their models produce. The habit of companies like Anthropic to release reports emphasizing the dangerousness of their products might be a decision that comes back to haunt them in future legal proceedings.

As a computer scientist and a digital ethicist, I'm both optimistic about the possibilities of A.I. and confounded by the terrifying and grim way that current technology leaders

insist on talking about it. This could have been a period of hopeful innovation, but instead our emotions are being manipulated by Silicon Valley's self-serving and morally untenable addiction to doom trolling. This communication strategy has to stop. The harm it's causing to the public's mental health has arguably outweighed the benefits that A.I. has so far delivered.

There have been some signs that this behavior is starting to shift. Perhaps in response to Anthropic's report on recursive self-improvement, OpenAI released a paper of its own, titled "Built to Benefit Everyone: Our Plan." It argues that "entirely automating everything is not the future we want" and that the company's goal is to produce a technology that makes people's lives strictly better, like the arrival of electric lighting in the early 20th century.

The paper reads like the standard overoptimistic, cheerleading marketing speak that we used to expect from major technology companies. Nothing about it aroused my emotions or stuck with me in any meaningful way. What a relief.

Cal Newport is a professor of computer science at Georgetown University and the author of "Deep Work."

*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.*