ABUNDANCE INSTITUTE

The AI Technopanic and Its Effects

A PRIMER

MAY 2024

→ Nirit Weiss-Blatt, PhD[1] → Adam Thierer[2] → Taylor Barkley[3]

The years 2023 and 2024 (so far) have been banner years for technopanics related to artificial intelligence (AI). The term *technopanic* refers to "intense public, political, and academic responses to the emergence or use of media or technologies."[4] This paper documents the most recent extreme rhetoric around

1 Communications Researcher with an expertise in tech journalism. Weiss-Blatt previously served as a Visiting Research Fellow at University of Southern California.

2 Innovation and Technology Policy Analyst, R Street Institute.

3 Director of Public Policy, Abundance Institute.

4 Adam Thierer, "Technopanics, Threat Inflation, and the Danger of an Information Technology Precautionary Principle," *Minnesota Journal of Law, Science and Technology* 14, no. 1 (2013): 311.

AI, identifies the incentives that motivate it, and explores the effects it has on public policy.

Granted, there are many serious issues related to the development of AI and the use of algorithmic and computational systems. Labeling the response to these innovations "technopanic" does not mean that the underlying technologies are harmless. Algorithmic systems can pose real threats. Many AI critics have raised concerns in a levelheaded fashion and have engaged in reasoned debate without resorting to rhetorical tactics such as fear appeals and threat inflation. Those tactics are meant to terrify the public and policymakers into taking extreme steps to curb or halt technological progress. Unfortunately, such scare tactics are becoming increasingly common today, and they often crowd out reasoned deliberation about the future of AI.

AI panic is often fueled by news media coverage. Looking at this coverage in light of the "technopanic" phenomenon prompts observers to notice what is being amplified and what is being ignored. These telling emphases and gaps in coverage are apparent in the case of the "existential risks" discourse that gives rise to some of the most extreme rhetoric and proposals in debates over the future of AI: Existential risk, or x-risk, gets most of the attention, while other risks are downplayed.

Our argument is that framing AI in such extremely negative terms can motivate policymakers to propose and adopt stringent rules that could chill or cancel beneficial innovation. Rhetoric has a cost. Thus, we outline a suite of proposals and recommendations for policymakers to adopt in order to provide the sober analysis required in their leadership positions. We also provide recommendations

for those in civil society and the media who wish humanity to reap as many benefits as possible from AI tools in the future.

Extreme Rhetoric on the Rise

In June 2022, a unique news story prompted the general public to recognize that large language models (LLMs) had dramatically improved. Google engineer Blake Lemoine argued that Google's LaMDA (Language Models for Dialogue Applications) is "sentient."[5] Among other claims, Lemoine said LaMDA resembles "an 8-year-old kid that happens to know physics."[6] The intense news cycle that followed this story prompted a worldwide discussion about the possibility of AI chatbots having self-awareness and feelings. The idea was received with skepticism. One *New York Times* article, for example, claimed that "robots can't think or feel, despite what the researchers who build them want to believe. A.I. is not sentient. Why do people say it is?"[7]

In August 2022, OpenAI gave one million people access to DALL-E 2. In November 2022, the company launched a user-friendly chatbot named ChatGPT. People started interacting with more advanced AI systems—"generative AI" tools—with Blake Lemoine's story in the background.[8]

5 Nirit Weiss-Blatt, "2023: The Year of AI Panic," *Techdirt*, December 22, 2023, https://www.techdirt.com/2023/12/22/2023-the-year-of-ai-panic/.

6 Quoted in Nitasha Tiku, "The Google Engineer Who Thinks the Company's AI Has Come to Life," *The Washington Post*, June 11, 2022, https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/.

7 Cade Metz, "A.I. Is Not Sentient. Why Do People Say It Is?," *The New York Times*, August 5, 2022, https://www.nytimes.com/2022/08/05/technology/ai-sentient-google.html.

8 Weiss-Blatt, "2023: The Year of AI Panic."

At first, news articles debated issues such as copyright and consent regarding AI-generated images (e.g., "AI Creating 'Art' Is an Ethical and Copyright Nightmare")[9] and how students will use ChatGPT to cheat on their assignments (e.g., "New York City Blocks Use of the ChatGPT Bot in Its Schools," "The College Essay Is Dead").[10] A turning point came after the release of *New York Times* columnist Kevin Roose's story on his disturbing conversation with Microsoft's new Bing chatbot.[11] It has since become known as the "Sydney tried to break up my marriage" story.[12] *The New York Times* cover page included parts of Roose's correspondence with the chatbot, headlined as "Bing's Chatbot Drew Me In and Creeped Me Out."[13] "The normal way that you deal with software that has a user interface bug is you just go fix the bug and apologize to the customer that triggered it," responded Kevin Scott, Microsoft's chief technology officer. "This one just happened to be one of the most-read stories in *The New York Times* history."[14]

9 Luke Plunkett, "AI Creating 'Art' Is an Ethical and Copyright Nightmare," *Kotaku*, August 25, 2022, https://kotaku.com/ai-art-dall-e-midjourney-stable-diffusion-copyright-1849388060.

10 Dan Rosenzweig-Ziff, "New York City Blocks Use of the ChatGPT Bot in Its Schools," *The Washington Post*, January 5, 2023, https://www.washingtonpost.com/education/2023/01/05/nyc-schools-ban-chatgpt/; Stephen Marche, "The College Essay Is Dead," *Atlantic*, December 6, 2022, https://www.theatlantic.com/technology/archive/2022/12/chatgpt-ai-writing-college-student-essays/672371/.

11 Kevin Roose, "A Conversation with Bing's Chatbot Left Me Deeply Unsettled," *The New York Times*, February 16, 2023, https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html. Material in this paragraph is adapted from Nirit Weiss-Blatt, "2023: The Year of AI Panic," *Techdirt*, December 22, 2023, https://www.techdirt.com/2023/12/22/2023-the-year-of-ai-panic/.

12 Roose, "Conversation with Bing's Chatbot."

13 Kevin Roose, "Bing's Chatbot Drew Me In and Creeped Me Out," *The New York Times*, February 17, 2023, cover page.

14 Quoted in Nilay Patel, "Microsoft CTO Kevin Scott Thinks Sydney Might Make a Comeback," *Verge*, May 23, 2023, https://www.theverge.com/23733388/microsoft-kevin-scott-open-ai-chat-gpt-bing-github-word-excel-outlook-copilots-sydney.

After that, things escalated quickly. The "existential risk" open letters appeared in spring 2023 (more on this later), and the "AI could kill everyone" scenario became a mainstream talking point. If it had found a platform only in the realm of British tabloids, it could have been dismissed as fringe sensationalism: "Humans 'Could Go Extinct' When Evil 'Superhuman' AI Robots Rise Up Like *The Terminator*."[15] But similar headlines spread across mass media and could soon be found even in prestigious news outlets (e.g., *The New York Times*: "If we don't master A.I., it will master us.").[16]

The demand for AI coverage produced full-blown exaggerations and clickbait metaphors. It snowballed into a competition of headlines. Patrick Grady and Daniel Castro from the Center for Data Innovation explain, "Once news media first get wind of a panic, it becomes a game of one-upmanship: the more outlandish the claims, the better."[17] This process reached its apogee with *Time* magazine's June 12, 2023, cover story on AI, teased with "THE END OF HUMANITY."[18]

The fact that grandiose claims such as the assertion that AI will cause human extinction have gained so much momentum is likely having distorting effects on public understanding, AI research

15 Brendan McFadden, "Humans 'Could Go Extinct' When Evil 'Superhuman' AI Robots Rise Up Like *The Terminator*," *Daily Star*, January 26, 2023, https://www.dailystar.co.uk/tech/news/humans-could-go-extinct-evil-29061844.

16 Nirit Weiss-Blatt, "The AI Doomers' Playbook," *Techdirt*, April 14, 2023, https://www.techdirt.com/2023/04/14/the-ai-doomers-playbook/.

17 Patrick Grady and Daniel Castro, "Tech Panics, Generative AI, and the Need for Regulatory Caution," Center for Data Innovation, May 1, 2023, https://datainnovation.org/2023/05/tech-panics-generative-ai-and-regulatory-caution/.

18 Katja Grace, "AI Is Not an Arms Race," *Time*, May 31, 2023, https://time.com/6283609/artificial-intelligence-race-existential-threat; *Time* cover vol. 201 no. 21, https://time.com/magazine/south-pacific/6284502/june-12th-2023-vol-201-no-21-asia-south-pacific/.

funding, corporate priorities, and government regulation.[19] This is why we present some of the more notable examples of essays and op-eds that promote the "existential risk" ideology and reflect the growing AI technopanic.

Here are some telltale signs that the AI technopanic mentality is coloring a specific piece of writing:

→ Quasi-religious rhetoric expressing fear of godlike powers of technology or suggesting that apocalyptic "end times" scenarios are approaching

→ The repeated use of dystopian pop culture allusions to frame discussions, such as references to The Terminator, The Matrix, or Black Mirror,[20] followed by implicit or explicit sympathy for violent actions or social uprisings to "stop the machine" or slow progress in some fashion

→ Calls for sweeping regulatory interventions to control technological progress, which may include widespread surveillance of research and development efforts or even militaristic interventions by governments, and possibly global government control

→ A tendency to ignore any trade-offs or downsides associated with these rhetorical ploys or the extreme recommendations set forth

19 Blake Richards et al., "The Illusion of AI's Existential Risk," Noema, July 18, 2023, https://www.noemamag.com/the-illusion-of-ais-existential-risk.

20 Adam Thierer, "How Science Fiction Dystopianism Shapes the Debate over AI & Robotics," *Discourse*, July 26, 2022, https://www.discoursemagazine.com/culture-and-society/2022/07/26/how-science-fiction-dystopianism-shapes-the-debate-over-ai-robotics/.

Overall, these articles have two commonalities: their focus on "inventing a monster and demanding that world leaders be as afraid of it as you are" and their promotion of dangerous ideas about how to tame it.[21]

Notable Examples of Extreme Rhetoric

Eliezer Yudkowsky, Cofounder of MIRI (the Machine Intelligence Research Institute)[22]

Repeatedly insisting that the world must "shut it all down," Yudkowsky says that stopping AI and computational science requires extreme interventions. In his preferred world, "allied nuclear countries are willing to run some risk of nuclear exchange if that's what it takes to reduce the risk of large AI training

21 Matthew Gault, "AI Theorist Says Nuclear War Preferable to Developing Advanced AI," *Vice*, March 31, 2023, https://www.vice.com/en/article/ak3dkj/ai-theorist-says-nuclear-war-preferable-to-developing-advanced-ai.

22 Yudkowsky published in *LessWrong* (an online forum he created) that MIRI's new strategy is "death with dignity" (Eliezer Yudkowsky, "MIRI Announces New 'Death with Dignity' Strategy," *LessWrong*, April 1, 2022, https://www.lesswrong.com/posts/j9Q8bRmwCgXRYAgcJ/miri-announces-new-death-with-dignity-strategy). He has estimated the chances of human survival to be 0%; thus, the probability of doom from AI is 100%. Since "survival is unattainable," he wrote, MIRI is shifting its focus to "helping humanity die with slightly more dignity." Yudkowsky published this article on April 1, so MIRI's communications lead, Rob Bensinger, explained in MIRI's newsletter that "although released on April Fools' Day (whence the silly title), the post body is an entirely non-joking account of Eliezer's current models, including his currently-high p(doom)" (Rob Bensinger, "July 2022 Newsletter," Machine Intelligence Research Institute, July 30, 2022, https://intelligence.org/2022/07/30/july-2022-newsletter/). Bensinger clarified in *LessWrong* that the post accurately "represents Eliezer's epistemic state": "The post is just honestly stating Eliezer's views, without any more hyperbole than a typical Eliezer post would have" (Rob Bensinger, April 6, 2022, comment on Yudkowsky, "MIRI Announces New 'Death with Dignity' Strategy," https://www.lesswrong.com/posts/j9Q8bRmwCgXRYAgcJ/miri-announces-new-death-with-dignity-strategy?commentId=FounAZsg4kFxBDiXs).

runs," he wrote in a *Time* essay.[23] He advocates several sweeping prohibitions:

> *"Shut down all the large GPU clusters (the large computer farms where the most powerful AIs are refined). Shut down all the large training runs. Put a ceiling on how much computing power anyone is allowed to use in training an AI system, and move it downward over the coming years to compensate for more efficient training algorithms—no exceptions for governments and militaries. Make immediate multinational agreements to prevent the prohibited activities from moving elsewhere. Track all GPUs sold. If intelligence says that a country outside the agreement is building a GPU cluster, be less scared of a shooting conflict between nations than of the moratorium being violated; be willing to destroy a rogue datacenter by airstrike."*[24]

Michael Cuenco, Associate Editor at American Affairs

Cuenco calls for "putting the AI revolution in a deep freeze" and stopping almost all digital innovation and computational progress. He advocates a literal "Butlerian Jihad," inspired by the *Dune* prequel of the same name, whose plot involves a conflict in which almost all computers, robots, and forms of AI are intentionally destroyed. Cuenco's call for policy action includes "a broader indefinite AI ban, accompanied by a social compact

23 Eliezer Yudkowsky, "Pausing AI Developments Isn't Enough. We Need to Shut It All Down," *Time*, March 29, 2023, https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/. Yudkowsky's radical suggestions were later described as a possible "effort to avoid annihilation at the hands of superintelligent A.I.": "Shutting it all down would call for draconian measures—perhaps even steps as extreme as those espoused by Yudkowsky, who recently wrote, in an editorial for *Time*, that we should 'be willing to destroy a rogue datacenter by airstrike,' even at the risk of sparking 'a full nuclear exchange.'" Matthew Hutson, "Can We Stop Runaway A.I.?," *The New Yorker*, May 16, 2023, https://www.newyorker.com/science/annals-of-artificial-intelligence/can-we-stop-the-singularity, quoting Yudkowsky, "Pausing AI Developments Isn't Enough."

24 Yudkowsky, "Pausing AI Developments Isn't Enough."

premised on the permanent prohibition on the use of advanced AI across multiple industries as a means of obtaining and preserving economic security." He says that any disruptive automation technology should be "subject to nationwide codes governing what is permissible or not in every industry. Any changes to these codes would have to be enacted at the national level, and the codes should, in practice, be politically difficult to loosen."[25]

Max Tegmark, President of the Future of Life Institute

Tegmark describes how scary it would be to lose control "to alien digital minds that don't care about humans":

> *"If superintelligence drives humanity extinct, it probably won't be because it turned evil or conscious, but because it turned competent, with goals misaligned with ours."*[26]

He concludes that in this scenario, "We get extincted as a banal side effect that we can't predict."[27]

25 Michael Cuenco, "We Must Declare Jihad against A.I.," *Compact*, April 28, 2023, https://compactmag.com/article/we-must-declare-jihad-against-a-i.

26 Max Tegmark, "The 'Don't Look Up' Thinking That Could Doom Us with AI," *Time*, April 25, 2023, https://time.com/6273743/thinking-that-could-doom-us-with-ai/. These types of statements resulted in newspapers' opinion sections being flooded with doomsday theories; one op-ed even suggested that "competing AGIs [artificial general intelligences] might use Earth's resources in ways incompatible with our survival. We could starve, boil or freeze." Zvi Mowshowitz, "AI Is the Most Dangerous Technology We've Ever Invented," *Telegraph*, May 9, 2023, https://www.telegraph.co.uk/news/2023/05/09/ai-is-the-most-dangerous-technology-weve-ever-invented.

27 Tegmark, "'Don't Look Up' Thinking."

Dan Hendrycks, Executive and Research Director at the Center for AI Safety

Hendrycks argues that:

> *"[E]volution tends to produce selfish behavior. Amoral competition among AIs may select for undesirable traits. Evolutionary pressure will likely ingrain AIs with behaviors that promote self-preservation. Humans are incentivized to cede more and more power to AI systems that cannot be reliably controlled, putting us on a pathway toward being supplanted as the earth's dominant species."*[28]

Yuval Harari, Professor at the Hebrew University of Jerusalem; Tristan Harris and Aza Raskin, Founders of the Center for Humane Technology

"We have summoned an alien intelligence," these authors argue. "We don't know much about it, except that it is extremely powerful and offers us bedazzling gifts but could also hack the foundations of our civilization." They worry that "A.I. could rapidly eat the whole of human culture" and that "soon we will also find ourselves living inside the hallucinations of nonhuman intelligence":

> *"We will finally come face to face with Descartes's demon, with Plato's cave, with the Buddhist Maya. A curtain of illusions could descend over the whole of humanity, and we might never again be able to tear that curtain away—or even realize it is there."*[29]

28 Dan Hendrycks, "The Darwinian Argument for Worrying about AI," *Time*, May 31, 2023, https://time.com/6283958/darwinian-argument-for-worrying-about-ai/.

29 Yuval Harari, Tristan Harris, and Aza Raskin, "You Can Have the Blue Pill or the Red Pill, and We're Out of Blue Pills," *The New York Times*, March 24, 2023, https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html.

Harari has called for stiff sanctions or even prison sentences for anyone who creates "fake people," although he has not defined what that means.[30]

Peggy Noonan, Opinion Columnist for *The Wall Street Journal*

In her first essay regarding AI, Noonan asks society to "pause it for a few years. Call in the world's counsel, get everyone in. Heck, hold a World Congress."[31] In a follow-up essay replete with religious metaphors, she warns that "developing AI is biting the apple. Something bad is going to happen. I believe those creating, fueling, and funding it want, possibly unconsciously, to be God and on some level think they are God." She also favorably cites Yudkowsky's *Time* essay.[32]

Erik Hoel, Assistant Professor at Tufts University

Hoel, who is a neuroscientist, writes that the time has come for panic and radical action against AI innovators.

> *"Panic is necessary because humans simply cannot address a species-level concern without getting worked up about it and catastrophizing," he claims. "We need to panic about AI and imagine the worst-case scenarios while, at the same time, occasionally admitting that we can pursue a politically-realistic AI safety agenda."*

30 Quoted in Hannah Devlin, "AI Firms Should Face Prison over Creation of Fake Humans, Says Yuval Noah Harari," *Guardian*, July 6, 2023, https://www.theguardian.com/technology/2023/jul/06/ai-firms-face-prison-creation-fake-humans-yuval-noah-harari.

31 Peggy Noonan, "A Six-Month AI Pause? No, Longer Is Needed," *The Wall Street Journal*, March 30, 2023, https://www.wsj.com/articles/a-six-month-ai-pause-no-longer-is-needed-civilization-danger-chat-gpt-chatbot-internet-big-tech-4b66da6e.

32 Peggy Noonan, "Artificial Intelligence in the Garden of Eden," *The Wall Street Journal*, April 20, 2023, https://www.wsj.com/articles/artificial-intelligence-in-the-garden-of-eden-adam-eve-gates-zuckerberg-technology-god-internet-40a4477a.

He fantasizes about "a civilization that pre-emptively stops progress on the technologies that threaten its survival" and rounds out his call to action by suggesting that anti-AI activists vandalize the Microsoft and OpenAI headquarters "because only panic, outrage, and attention lead to global collective action."[33]

Steve Rose, Assistant Features Editor at the *Guardian*

Rose has collected five essays on "the ways AI might destroy the world." Max Tegmark's essay compares future human extinction with recent extinctions, such as that of the West African black rhinoceros and orangutans in Borneo. The essay by Ajeya Cotra, who oversee Open Philanthropy's "Potential risks from advanced artificial intelligence" program, compares GPT-4's "brain" to a squirrel's brain and recommends that the technology ratchet up to a hedgehog brain and not advance to the equivalent of a human brain. (That's a lot of animals in one article about AI!) Yoshua Bengio discusses the survival instinct:

> *"When we create an entity that has survival instinct, it's like we have created a new species. Once these AI systems have a survival instinct, they might do things that can be dangerous for us."*

Eliezer Yudkowsky suggests what such a superintelligence would do: (1) It is "probably going to want to do things that kill us as a side-effect, such as building so many power plants that run off nuclear fusion—because there is plenty of hydrogen in the oceans—that the oceans boil." (2) "It could build itself a tiny molecular laboratory and manufacture and release lethal bacteria.

33 Erik Hoel, "'I Am Bing, and I Am Evil,'" *Intrinsic Perspective*, February 16, 2023, https://www.theintrinsicperspective.com/p/i-am-bing-and-i-am-evil.

What that looks like is everybody on Earth falling over dead inside the same second."[34]

Open Letters and the Escalation of the Technopanic

Extreme rhetoric can also be found in open letters, which draw considerable media coverage.[35] See table 1 for details about the two open letters regarding x-risk released in 2023.

TABLE 1 | Basic Details about the Two Existential Risk Open Letters

LETTER	FIRST OPEN LETTER: SIX-MONTH PAUSE	SECOND OPEN LETTER: AI AS RISKY AS PANDEMICS AND NUCLEAR WAR
DATE	March 22, 2023	May 30, 2023
PUBLISHER	Future of Life Institute, founded by Jaan Tallinn, Max Tegmark, Victoria Krakovna, Anthony Aguirre, and Meia Chita-Tegmark	Center for AI Safety, founded by Dan Hendrycks and Oliver Zhang
FUNDING	Until 2021, the institute was primary funded by Elon Musk; then, Vitalik Buterin donated $665.8 million through a Shiba Inu memecoin	The center's primary funder in 2022 was Open Philanthropy, an effective altruism grant-making organization run by Dustin Moskovitz and Cari Tuna

Sources: Brendan Bordelon, "The little-known AI group that got $660 million," *Politico*, March 26, 2024, https://www.politico.com/news/2024/03/25/a-665m-crypto-war-chest-roils-ai-safety-fight-00148621; "Center for AI Safety—General Support (2022)," Open Philanthropy, accessed January 16, 2024, https://www.openphilanthropy.org/grants/center-for-ai-safety-general-support/; Brendan Bordelon, "AI Doomsayers Funded by Billionaires Ramp Up Lobbying," *Politico*, February 23, 2024, https://www.politico.com/news/2024/02/23/ai-safety-washington-lobbying-00142783.

34 Steve Rose, "Five Ways AI Might Destroy the World: 'Everyone on Earth Could Fall Over Dead in the Same Second,'" *The Guardian*, July 7, 2023, https://www.theguardian.com/technology/2023/jul/07/five-ways-ai-might-destroy-the-world-everyone-on-earth-could-fall-over-dead-in-the-same-second.

35 Nirit Weiss-Blatt (@DrTechlash), "Why do Doomers initiate Open Letters? Because they draw all the attention. How well does this PR stunt work? See for yourself," Twitter, June 3, 2023, 6:53 p.m., https://twitter.com/DrTechlash/status/1665129656683761664.

The first notable open letter, initiated by the Future of Life Institute (FLI), was released on March 22, 2023. FLI, as described on the Effective Altruism Forum, is "a non-profit that works to reduce existential risk from powerful technologies, particularly artificial intelligence."[36] In this widely discussed letter, various individuals called for AI labs "to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4."[37] The letter argued that "if such a pause cannot be enacted quickly, governments should step in and institute a moratorium." It warned of "an out-of-control race to develop and deploy ever more powerful digital minds" and "catastrophic effects on society."[38] The reasoning behind the immediate pause was expressed in the form of a rhetorical question: "*Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us?"[39]

Apparently, these speculative assumptions weren't enough, because the cofounder of the Future of Life Institute, Max Tegmark, added in an interview,

> *We just had a little baby, and I keep asking myself . . . How old is he even gonna get? There's a pretty large chance we're not gonna make it as humans. There won't be any humans on the planet in the not-too-distant future. This is the kind of cancer which kills all of humanity.*[40]

36 Future of Life Institute, *Effective Altruism Forum*, *https://forum.effectivealtruism.org/topics/future-of-life-institute*.

37 Future of Life Institute, "Pause Giant AI Experiments: An Open Letter," March 22, 2023, https://futureoflife.org/open-letter/pause-giant-ai-experiments.

38 Future of Life Institute, "Pause Giant AI Experiments."

39 Future of Life Institute, "Pause Giant AI Experiments."

40 Max Tegmark, interview by Lex Fridman, "Max Tegmark: The Case for Halting AI Development," *Lex Fridman Podcast*, YouTube video, 2:48:12, April 13, 2023, https://youtu.be/VcVfceTsD0A.

The second x-risk open letter, initiated by the Center for AI Safety, was released on May 30, 2023. It raised the rhetorical panic level by publishing a single statement: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[41] This letter was launched in *The New York Times* with the headline, "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn."[42] Consequently, Robert Wiblin, the former executive director of the Centre for Effective Altruism and the current director of research at 80,000 Hours, declared that "AI extinction fears have largely won the public debate."[43] Max Tegmark celebrated how the "AI extinction threat is going mainstream."[44]

In the aftermath of these dramatic warnings, some AI pioneers have escalated their rhetoric. Geoffrey Hinton (one of the signers of the second letter) said, "I think it's quite conceivable that humanity is just a passing phase in the evolution of intelligence."[45] "The alarm bell I'm ringing has to do with the existential threat of [powerful AI systems] taking control."[46] "It's not just science

41 Center for AI Safety, "Statement on AI Risk," released May 30, 2023 (undated), https://www.safe.ai/statement-on-ai-risk.

42 Quoted in Kevin Roose, "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn," *The New York Times*, May 30, 2023, https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html.

43 Robert Wiblin (@robertwiblin), "AI extinction fears have largely won the public debate. Staff at AI labs, governments and the public are all very worried and willing to take action," Twitter, May 30, 2023, 8:26 a.m., https://twitter.com/robertwiblin/status/1663522331954761728.

44 Max Tegmark, interview by Krishnan Guru-Murthy, "AI Extinction Threat Is 'Going Mainstream,' Says Max Tegmark," *Channel 4*, video, 8:12, May 30, 2023, https://www.channel4.com/news/ai-extinction-threat-is-going-mainstream-says-max-tegmark.

45 Geoffrey Hinton, interview by Will Douglas Heaven, "Video: Geoffrey Hinton Talks about the 'Existential Threat' of AI," *MIT Technology Review*, video, 38:26, May 3, 2023, https://www.technologyreview.com/2023/05/03/1072589/video-geoffrey-hinton-google-ai-risk-ethics.

46 Quoted in Craig S. Smith, "Geoff Hinton, AI's Most Famous Researcher, Warns of 'Existential Threat' from AI," *Forbes*, May 4, 2023, https://www.forbes.com/sites/craigsmith/2023/05/04/geoff-hinton-ais-most-famous-researcher-warns-of-existential-threat.

fiction. It's not just fear-mongering. It is a real risk we need to think about."[47] Another pioneer, Yoshua Bengio (who signed both letters), shared that he believes that the technologies have become so capable that they risk triggering a world-ending catastrophe, whether as rogue sentient entities or in the hands of a human. "If it's an existential risk, we may have one chance, and that's it."[48] Jaan Tallinn, cofounder of the Future of Life Institute and the Centre for the Study of Existential Risk and biggest donor to the Survival and Flourishing Fund, said in an interview, "I've not met anyone in AI labs who says the risk is less than 1% of blowing up the planet. It's important that people know lives are being risked."[49]

CEOs of AI start-ups have begun emphasizing similar existential risk scenarios. Emad Mostaque, CEO of Stability AI, signed both x-risk open letters. He said, "The worst case scenario is that [AI] proliferates and basically it controls humanity."[50] He also explained on Twitter, "There's so many ways to wipe out humanity for something that can be more persuasive than anyone & replicate itself & gather any resources."[51] (It is actually "AI apocalypse"

47 Quoted in Aaron Anandji, "AI Risks: Overblown, Existential Threat, or Something Else? VCs and Tech Experts Disagree," *BetaKit*, July 6, 2023, https://betakit.com/ai-risks-overblown-existential-threat-or-something-else-vcs-and-tech-experts-disagree.

48 Quoted in Matteo Wong, "AI Doomerism Is a Decoy," *The Atlantic*, June 2, 2023, https://www.theatlantic.com/technology/archive/2023/06/ai-regulation-sam-altman-bill-gates/674278.

49 Quoted in Liron Shapira (@liron), "Jaan Tallinn, lead investor of @AnthropicAI's $124M Series A, said today: 'I've not met anyone in AI labs who says the risk . . . ," Twitter, May 12, 2023, 3:50 a.m., https://twitter.com/liron/status/1656929936639430657.

50 Quoted in Laura Kuenssberg, "AI Creator on the Risks, Opportunities and How It May Make Humans 'Boring,'" *BBC News*, May 13, 2023, https://www.bbc.com/news/uk-politics-65582386.

51 Emad Mostaque (@EMostaque), "I don't get folk who say there is no existential risk from AGI. Do y'all have no imagination? There's so many ways to wipe out humanity," Twitter, May 20, 2023, 5:46 p.m., https://twitter.com/EMostaque/status/1660039251835117568.

scenarios that replicate and gather resources.) When Sam Altman, OpenAI's CEO, shared his worst-case scenario of AI, it was "lights out for all of us."[52] In his interview tour, he frequently emphasized that he is "super-nervous,"[53] that he empathizes "with people who are a lot afraid,"[54] and that "there is a legitimate existential risk here."[55] Altman also signed the second open letter, which compared the risk from AI to the risk from nuclear war and pandemics.

Potential Incentives for Extreme Rhetoric

Why would someone frame their own company the way Sam Altman has? Making one's products "the most important—and hopeful, and scary—project in human history"[56] is part of the marketing strategy: "The paranoia is the marketing."[57] "If you want people to think what you're working on is powerful, it's a good idea to make them *fear it*," explains François Chollet, an AI researcher at Google.[58]

52 Quoted in Connie Loizos, "Video Is Coming and More from OpenAI CEO Sam Altman," *TechCrunch*, January 18, 2023, https://techcrunch.com/2023/01/17/that-microsoft-deal-isnt-exclusive-video-is-coming-and-more-from-openai-ceo-sam-altman.

53 "Sam Altman on What Makes Him 'Super Nervous' about AI," On with Kara Swisher, *Intelligencer* (*New York* Magazine), March 23, 2023, https://nymag.com/intelligencer/2023/03/on-with-kara-swisher-sam-altman-on-the-ai-revolution.html.

54 Sam Altman, interview by Lex Fridman, "Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI," *Lex Fridman Podcast*, YouTube video, 2:23:56, March 25, 2023, https://youtu.be/L_Guz73e6fw.

55 Quoted in Josiah Mackenzie, "OpenAI (ChatGPT) CEO Sam Altman on AI and the Future of Technology," *Hospitality Daily*, May 26, 2023, https://www.hospitalitynet.org/opinion/4116587.html.

56 Sam Altman, "Planning for AGI and beyond," *OpenAI Blog*, February 24, 2023, https://openai.com/blog/planning-for-agi-and-beyond.

57 Nirit Weiss-Blatt (@DrTechlash), "'The paranoia is the marketing,'" Twitter, April 2, 2023, 1:06 a.m., https://twitter.com/DrTechlash/status/1642755460083355648.

58 Quoted in Will Douglas Heaven, "How Existential Risk Became the Biggest Meme in AI," *MIT Technology Review*, June 19, 2023, https://www.technologyreview.com/2023/06/19/1075140/how-existential-risk-became-biggest-meme-in-ai, emphasis added.

"AI doomsaying is absolutely everywhere right now," wrote Brian Merchant, the *Los Angeles Times* tech columnist. "Which is exactly the way that OpenAI, the company that stands to benefit the most from everyone believing its product has the power to remake—or unmake—the world, wants it." Merchant explains Altman's science-fiction-infused marketing frenzy: "Scaring off customers isn't a concern when what you're selling is the fearsome power that your service promises."[59]

One of us (Nirit Weiss-Blatt) has published a guide to the "AI existential risk" ecosystem. Weiss-Blatt classifies the AI panic facilitators as adopting either a "Panic-as-a-Business" attitude or an "AI Panic Marketing" attitude:[60] In "Panic-as-a-Business," the panic promoters[61] are basically saying, "We believe humans will be wiped out by a Godlike, superintelligent AI. All resources should be focused on that!"[62]

→ In "AI Panic Marketing," the panic promoters are basically saying, "We're building a powerful, godlike, superintelligent AI. See how much is invested in taming it!"

59 Brian Merchant, "Column: Afraid of AI? The Startups Selling It Want You to Be," *Los Angeles Times*, March 31, 2023, https://www.latimes.com/business/technology/story/2023-03-31/column-afraid-of-ai-the-startups-selling-it-want-you-to-be.

60 Nirit Weiss-Blatt, "The AI Panic Campaign—Part 2," *AI Panic*, October 15, 2023, *https://www.aipanic.news/p/the-ai-panic-campaign-part-2*; Nirit Weiss-Blatt, "Ultimate Guide to 'AI Existential Risk' Ecosystem," *AI Panic*, December 5, 2023, *https://www.aipanic.news/p/ultimate-guide-to-ai-existential*.

61 See Nirit Weiss-Blatt (@DrTechlash), "'A Taxonomy of AI Panic Facilitators': A visualization of leading AI Doomers (X-risk open letters, media interviews & OpEds)," Twitter, June 1, 2023, 10:51 a.m., https://twitter.com/DrTechlash/status/1675155157880016898; AlgorithmWatch (@algorithmwatch), "Mirror, Mirror on the wall, who is the biggest Panic-creator of them all? Inspired by @DrTechlash, check out our Taxonomy of AI Panic Facilitators," Twitter, July 11, 2023, 12:15 p.m., https://twitter.com/algorithmwatch/status/1678800286062727168.

62 Nitasha Tiku, "How Elite Schools Like Stanford Became Fixated on the AI Apocalypse," *The Washington Post*, July 5, 2023, https://www.washingtonpost.com/technology/2023/07/05/ai-apocalypse-college-students.

A *New York Times* article profiling the AI company Anthropic, titled "Inside the White-Hot Center of A.I. Doomerism," demonstrates the "AI Panic Marketing" attitude. Because of the company's "effective altruism" culture, its employees shared a prediction that there was a "20 percent chance of imminent doom." "They worry, obsessively, about what will happen if A.I. alignment isn't solved by the time more powerful A.I. systems arrive," observes the article's author, Kevin Roose. However, thanks to Anthropic's unique "Constitutional A.I." technique, "you get an A.I. system that largely polices itself and misbehaves less frequently than chatbots trained using other methods," the company claimed.[63] *The New York Times* published Anthropic's profile the day the company launched its new chatbot, "Claude 2."

In July 2023, OpenAI launched a "Superalignment" team to control "superintelligence."[64] The team's opening statement is another example of extreme rhetoric and industry motivation: "Superintelligence will be the most impactful technology humanity has ever invented, and could help us solve many of the world's most important problems. But the vast power of superintelligence could also be very dangerous, and could lead to the disempowerment of humanity or *even human extinction*."[65] The solution? The company pledged to dedicate 20 percent of its computational power to

63 Kevin Roose, "Inside the White-Hot Center of A.I. Doomerism," *The New York Times*, July 11, 2023, https://www.nytimes.com/2023/07/11/technology/anthropic-ai-claude-chatbot.html.

64 Kai Xiang Teo, "OpenAI Is So Worried about AI Causing Human Extinction, It's Putting Together a Team to Control 'Superintelligence,'" *Business Insider*, July 7, 2023, https://www.businessinsider.in/tech/news/openai-is-so-worried-about-ai-causing-human-extinction-its-putting-together-a-team-to-control-superintelligence/articleshow/101566624.cms.

65 Jan Leike and Ilya Sutskever, "Introducing Superalignment," OpenAI, July 5, 2023, https://openai.com/blog/introducing-superalignment, emphasis added.

"solving the problem." (One Superalignment team member called his team the "notkilleveryoneism" team.)[66]

AI ethicist Rumman Chowdhury characterized this attitude as a "disempowerment narrative":

> *The general premise of all of this language is, "We have not yet built but will build a technology that is so horrible that it can kill us. But clearly, the only people skilled to address this work are us, the very people who have built it, or who will build it." That is insane.*[67]

The tech sector has a tendency to "overstate the capabilities of their products," according to Will Douglas Heaven, *MIT Technology Review*'s senior editor for AI. Heaven suggested that "if something sounds like bad science fiction, maybe it is."[68] Emily Bender, a University of Washington professor, would probably agree: "The whole thing looks to me like a media stunt, to try to grab the attention of the media, the public, and policymakers and focus everyone on the distraction of sci-fi scenarios," she said. "This would seem to serve two purposes: it paints their tech as way more powerful and effective than it is, and it takes the focus away from the actual harms being done, now."[69]

66 Nat McAleese (@__nmca__), "1) Yes, this is the notkilleveryoneism team," Twitter, July 4, 2023, 1:19 p.m., https://twitter.com/__nmca__/status/1676641876537999385.

67 Quoted in Lorena O'Neil, "These Women Tried to Warn Us about AI," *Rolling Stone*, August 12, 2023, https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/.

68 Quoted in Melissa Heikkilä, "How to Talk about AI (Even If You Don't Know Much about AI)," *MIT Technology Review*, May 30, 2023, https://www.technologyreview.com/2023/05/30/1073680/how-to-talk-about-ai-even-if-you-dont-know-much-about-ai/.

69 Quoted in Chloe Xiang, "AI CEOs Say AI Poses 'Risk of Extinction,' Are Trying to Find the Guy Who Did This," *Vice*, May 30, 2023, https://www.vice.com/en/article/xgwy94/ai-ceos-say-ai-poses-risk-of-extinction-are-trying-to-find the-guy-who-did-this.

Kyunghyun Cho, a prominent AI researcher from New York University, explained the "AI Panic Marketing" attitude as a "savior complex": "They all want to save us from the inevitable doom that only they see and think only they can solve. These people are loud, but they're still a fringe group within the whole society, not to mention the whole machine learning community."[70]

This brings us to the "Panic-as-a-Business" attitude, which has been adopted by effective altruism. This movement created a network comprising hundreds of organizations that are led by a relatively small group of influential leaders and a handful of key organizations (e.g., Open Philanthropy).[71] In recent years, these effective altruism institutes[72] have promoted the "AGI [Artificial General Intelligence] apocalypse" ideology through field-building of "AI safety" and "AI alignment" research (which aim to align future AI systems with human values).[73]

Ever since Eliezer Yudkowsky, one of the founders of the field of "AI alignment," published his influential *Time* op-ed, he has been on a media blitz. "I expected to be a tiny voice shouting into

70 Quoted in Sharon Goldman, "Top AI Researcher Dismisses AI 'Extinction' Fears," *VentureBeat*, June 1, 2023, https://venturebeat.com/ai/top-ai-researcher-dismisses-ai-extinction-fears-challenges-hero-scientist-narrative/.

71 Mollie Gleiberman, "Effective Altruism and the strategic ambiguity of 'doing good,'" *University of Antwerp*, January 2023, *https://repository.uantwerpen.be/docman/irua/371b9dmotoM74*. See "Appendix B: EA Organizations."

72 Ellen Huet, "The Real-Life Consequences of Silicon Valley's AI Obsession," *Bloomberg*, March 7, 2023, https://www.bloomberg.com/news/features/2023-03-07/effective-altruism-s-problems-go-beyond-sam-bankman-fried#xj4y7vzkg; Timnit Gebru, "Effective Altruism Is Pushing a Dangerous Brand of 'AI Safety,'" *Wired*, November 30, 2022, https://www.wired.com/story/effective-altruism-artificial-intelligence-sam-bankman-fried.

73 Nirit Weiss-Blatt, "Effective Altruism Funded the 'AI Existential Risk' Ecosystem with Half a Billion Dollars," *AI Panic*, December 5, 2023, *https://www.aipanic.news/p/effective-altruism-funded-the-ai*; Machine Intelligence Research Institute, "Why AI Safety?," accessed January 22, 2024, *https://intelligence.org/why-ai-safety/*.

the void, and people listened instead," Yudkowsky admitted. "So, I doubled down on that," referring to his AI doom scenarios.[74]

Among the effective altruism movement's leading donors are Jaan Tallinn (cofounder of the Future of Life Institute), Vitalik Buterin (cofounder of Ethereum), Sam Bankman-Fried (disgraced founder of FTX), and Elon Musk (CEO of Tesla and xAI). A notable institution in this realm is Open Philanthropy (cofounded by Dustin Moskovitz), which has funneled nearly half a billion dollars into developing a pipeline of talent to fight "rogue AI." According to *The Washington Post*, it included building a scaffolding of think tanks, YouTube channels, prize competitions, grants, research funding, and scholarships.[75] The interest in working on AI x-risk did not arise organically, as Princeton computer science PhD candidate Sayash Kapoor points out: "It has been very strategically funded by organizations that make x-risk a top area of focus."[76]

A major component of effective altruism's philosophy is "longtermism"—a focus on the distant future's potential catastrophes. Elon Musk called longtermism a "close match" to his own ideology.[77] After attending the "AI Forum" called by the US Congress in September 2023, Musk told reporters that, though

74 Will Henshall, "Eliezer Yudkowsky," *Time*, September 7, 2023, https://time.com/collection/time100-ai/6309037/eliezer-yudkowsky/.

75 Nitasha Tiku, "How Elite Schools Like Stanford Became Fixated on the AI Apocalypse." July 5, 2023, https://www.washingtonpost.com/technology/2023/07/05/ai-apocalypse-college-students/.

76 Quoted in Louise Matsakis, "The Princeton Researchers Calling Out 'AI Snake Oil,'" Semafor, September 15, 2023, *https://www.semafor.com/article/09/15/2023/the-princeton-researchers-calling-out-ai-snake-oil*.

77 Elon Musk (@elonmusk), "Worth reading. This is a close match for my philosophy," Twitter, August 1, 2022, 1:15 a.m., https://twitter.com/elonmusk/status/1554335028313718784. See also Émile P. Torres, "The Dangerous Ideas of 'Longtermism' and 'Existential Risk,'" *Current Affairs*, July 28, 2021, *https://www.currentaffairs.org/2021/07/the-dangerous-ideas-of-longtermism-and-existential-risk*.

the chances are low, there's a nonzero possibility that "AI will kill us all."[78] According to Reid Hoffman (who founded OpenAI with him), Musk's "whole approach to AI is: AI can only be saved if I deliver, if I build it." Five years after Musk left OpenAI, Sam Altman made a similar observation: "Elon desperately wants the world to be saved. But only if he can be the one to save it."[79]

The two x-risk open letters cited above spurred the AI panic as they gathered tens of thousands of signatories. These signatories provided credibility that drove the letters to fame. While people like Max Tegmark were interviewed about why they signed the letters, most signatories were "faceless," and their motivations were unclear. That is why the paper "Why They're Worried" is an interesting attempt to understand the views of those who signed the first open letter. The researchers interviewed early signatories about "how their beliefs relate to the letter's stated goals."[80] The signatories' answers revealed that their concerns were not centered on "human extinction" at all.

Most of the interviewed signatories indicated that they did not "envision the apocalyptic scenario that some parts of the document warn about." For example, Moshe Vardi "disagreed with almost every line"; Ricardo Baeza-Yates "thought that the request was not the right one and also that the reasons were the wrong ones"; an anonymous signatory "didn't read it all and [doesn't] buy into it all." The researchers concluded that "while a few aligned

78 Quoted in Antonio Pequeño IV, "AI Forum: Tech Executives Warn of AI Dangers and 'Superintelligence' in Closed-Door Meeting," *Forbes*, September 13, 2023, https://www.forbes.com/sites/antoniopequenoiv/2023/09/13/ai-forum-tech-executives-warn-of-ai-dangers-and-superintelligence-in-closed-door-meeting/.

79 Quoted in Ronan Farrow, "Elon Musk's Shadow Rule," *The New Yorker*, August 21, 2023, https://www.newyorker.com/magazine/2023/08/28/elon-musks-shadow-rule.

80 Isabella Struckman and Sofie Kupiec, "Why They're Worried: Examining Experts' Motivations for Signing the 'Pause Letter,'" *arXiv*, June 19, 2023, 5, https://arxiv.org/abs/2306.00891.

with the letter's existential focus, many . . . were far more preoccupied with problems relevant to today."[81] Nonetheless, as Nirit Weiss-Blatt has observed elsewhere, "They lent their name to the extreme AI doomers."[82]

What this "six-month pause" letter did was to normalize "expressing deep AI fears."[83] Looking back on its impact, Max Tegmark shared, "I was overwhelmed by the success of the letter in bringing about this sorely needed conversation. It was amazing how it exploded into the public sphere." Part of this "success" is that now "all sides realize that if anyone builds out of control superintelligence, we all go extinct."[84]

The result is that, unlike the "techlash" against social media, the current "AI techlash" amplifies the x-risk angle. "This is historically quite abnormal," said Kevin Roose, tech columnist at *The New York Times*.[85] Amid all the previous criticism of Facebook regarding political polarization, disinformation, and kids' mental health, creator Mark Zuckerberg wasn't blamed for wiping out humanity (nor did he warn that his products might do so). The AI techlash feels overwhelming and unprecedented—because it is.

81 Struckman and Kupiec, "Why They're Worried," 10.

82 Quoted in Will Knight, "A Letter Prompted Talk of AI Doomsday. Many Who Signed Weren't Actually AI Doomers," *Wired*, August 17, 2023, https://www.wired.com/story/letter-prompted-talk-of-ai-doomsday-many-who-signed-werent-actually-doomers/.

83 Ryan Heath, "The Great AI 'Pause' That Wasn't," *Axios*, September 22, 2023, https://www.axios.com/2023/09/22/ai-letter-six-month-pause.

84 Quoted in Reed Albergotti, "Author of 'Pause AI' Letter Reflects on Its Impact," *Semafor*, September 22, 2023, https://www.semafor.com/article/09/22/2023/author-of-pause-ai-letter-reflects-on-its-impact.

85 Quoted in Nirit Weiss-Blatt (@DrTechlash), "@kevinroose about AI Doomers: 'I've been covering tech for more than a decade. I have to say this is highly unusual,'" Twitter, June 1, 2023, 12:33 a.m., https://twitter.com/DrTechlash/status/1664127910431846400.

The Media's Incentives and Role in Fueling Doomsaying

There are "Top 10 AI Frames" that encapsulate the media's "know-how" for covering AI.[86] These AI descriptions are organized from the most positive to the most negative. Since the media is drawn to extreme depictions, the AI coverage includes mainly exaggerated utopian scenarios (on how AI will save humanity) alongside exaggerated dystopian scenarios (on how AI will destroy humanity). Recently, the most negative frame, the "existential threat" theme, has been getting the most attention.[87]

Ian Hogarth, author of the column "We Must Slow Down the Race to God-Like AI," shared that this column was "the most read story" in the *Financial Times* the day it was published.[88] Similarly, Steve Rose, assistant features editor for *The Guardian*, shared this simple truth: "So far, 'AI worst case scenarios' has had 5 x as many readers as 'AI best case scenarios.'"[89] Hogarth's *Financial Times* op-ed stated that "God-like AI . . . could usher in the obsolescence or destruction of the human race."[90] *The Guardian* declared in a headline that "Everyone on Earth Could Fall Over Dead in the Same Second."[91]

86 Nirit Weiss-Blatt, "Your Guide to the Top 10 'AI Media Frames,'" *AI Panic*, September 10, 2023, https://www.aipanic.news/p/your-guide-to-the-top-10-ai-media.

87 Nirit Weiss-Blatt, "Overwhelmed by All the Generative AI Headlines? This Guide Is for You," *Techdirt*, March 1, 2023, https://www.techdirt.com/2023/03/01/overwhelmed-by-all-the-generative-ai-headlines-this-guide-is-for-you.

88 Ian Hogarth (@soundboy), "Cool to wake up and find out it was the most read story on FT.com yesterday! Great to see so many people engaging and sending helpful ideas and feedback," Twitter, April 14, 2023, 3:53 a.m., https://twitter.com/soundboy/status/1646783609603338240.

89 Steve Rose (@steverose7), Twitter, July 7, 2023, 12:27 p.m., https://twitter.com/steverose7/status/1677353629634764800.

90 Ian Hogarth, "We Must Slow Down the Race to God-Like AI," *Financial Times*, April 13, 2023, https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2.

91 Rose, "Five Ways AI Might Destroy the World."

It's not surprising that these articles were successful. Tragedy and catastrophe garner attention. After all, according to the journalistic marketing truism, "If it bleeds, it leads."

According to Paris Martineau, a tech reporter at *The Information*, who was interviewed by *Columbia Journalism Review*, we need to consider the structural headwinds buffeting journalism—the collapse of advertising revenue, shrinking editorial budgets, smaller newsrooms, and the demand for SEO traffic. In a perfect world, all reporters would have the time and resources to write ethically framed, non-science-fiction-like stories about AI. But they do not. "It is systemic," Martineau said.[92]

92 Quoted in Jem Bartholomew and Dhrumil Mehta, "How the Media Is Covering ChatGPT," *Columbia Journalism Review*, May 26, 2023, https://www.cjr.org/tow_center/media-coverage-chatgpt.php; also appeared in Nirit Weiss-Blatt, "What's Wrong with AI Media Coverage & How to Fix It," *AI Panic*, September 10, 2023, https://www.aipanic.news/p/whats-wrong-with-ai-media-coverage.

Since the media plays a crucial role in the self-reinforcing cycle of AI doomerism, Nirit Weiss-Blatt has outlined seven ways AI media coverage fails us, using the acronym "AI PANIC":[93]

A I HYPE AND CRITI-HYPE
- → AI hype describes when overconfident techies brag about their AI systems (also termed AI boosterism).
- → AI criti-hype describes when overconfident doomsayers accuse those AI systems of atrocities (also termed AI doomerism).
- → Both overpromise the technology's capabilities.

I NDUCING SIMPLISTIC, BINARY THINKING
- → Discussion is either simplistically optimistic or simplistically pessimistic.
- → When companies' founders are referred to as "charismatic leaders," AI ethics experts as "critics" or "skeptics," and doomsayers (without expertise in AI) as "AI experts," this distorts how the public perceives, understands, and participates in these discussions.

P ACK JOURNALISM
- → Pack journalism encourages copycat behavior: different news outlets report the same story from the same perspective.
- → It leads to media storms.
- → In the current media storm, AI doomers' fearmongering overshadows the real consequences of AI. The resulting conversation is not productive, yet the press runs with it.

A NTHROPOMORPHIZING AI
- → Attributing human characteristics to AI misleads people.
- → Anthropomorphizing begins with words like intelligence and learning and moves on to consciousness and sentience, as if the machine has experiences, emotions, opinions, or motivations. AI is not a human being.

N ARROW FOCUS ON THE EDGES OF THE DEBATE
- → The selection of topics for attention and the framing of these topics are powerful agenda-setting roles.
- → This is why it's unfortunate that the loudest shouters lead the AI discussion's framing.

I NTERCHANGING QUESTION MARKS AND EXCLAMATION POINTS
- → Sensational, deterministic headlines prevail over nuanced discussions.
- → "Artificial General Intelligence Will Destroy Us!" and "Artificial General Intelligence Will Save Us!" make for good headlines, not good journalism.

C ONVERSING SCI-FI SCENARIOS AS CREDIBLE PREDICTIONS
- → "AI will get out of control and kill everyone." This scenario doesn't need any proof or factual explanation.
- → We saw it in Hollywood movies! So it must be true . . . right?

93 The following list originally appeared in Weiss-Blatt, "What's Wrong with AI Media Coverage & How to Fix It."

"The proliferation of sensationalist narratives surrounding artificial intelligence—fueled by interest, ignorance, and opportunism—threatens to derail essential discussions on AI governance and responsible implementation," warn Divyansh Kaushik and Matt Korda from the Federation of American Scientists.[94] The next section will show how it has already derailed the AI governance discussion.

Effect on Politicians

Politicians are paying attention to the AI panic. According to the National Conference of State Legislatures, over 90 AI-related bills had been introduced by midsummer 2023, many of them pushing for extensive regulation.[95] As of April 2024 that number had increased to nearly 600 bills in the states and nearly 100 bills in Congress.[96] The attention is likely to increase even more. "Regulators around the world are now scrambling to decide how to regulate the technology, while respected researchers are warning of longer-term harms, including that the tech might one day surpass human intelligence," wrote Gerrit De Vynck in *The Washington Post*. "There's an AI-focused hearing on Capitol Hill nearly every week."[97]

94 Divyansh Kaushik and Matt Korda, "Panic about Overhyped AI Risk Could Lead to the Wrong Kind of Regulation," *Vox*, July 3, 2023, https://www.vox.com/future-perfect/2023/7/3/23779794/artificial-intelligence-regulation-ai-risk-congress-sam-altman-chatgpt-openai.

95 National Conference of State Legislatures. "Artificial Intelligence 2023 Legislation." Last updated January 12, 2024. https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation.

96 Multistate.ai. "Artificial Intelligence (AI) Legislation." Accessed April 23, 2024. https://www.multistate.ai/artificial-intelligence-ai-legislation.

97 Gerrit De Vynck, "Google's AI Ambassador Walks a Fine Line between Hype and Doom," The *Washington Post*, August 9, 2023, https://www.washingtonpost.com/technology/2023/08/09/google-james-manyika-ai-existential-threat/.

At the federal level, Congress, regulatory agencies, and the White House are all reacting to the public discourse by releasing guidance documents, memos, and op-eds. In May 2023, the White House hosted a summit of many of the leading generative AI CEOs. At one Senate Judiciary Committee hearing in May 2023, Sen. John Kennedy suggested that political deliberations about these issues should begin with the assumption that AI wants to kill us.[98] The State Department spent $250,000 in November 2022 to commission a report released in February 2024 that compared advanced AI models to weapons of mass destruction. The report included recommendations to create a new federal regulatory agency, an international AI agency, and for Congress to outlaw "AI models using more than a certain level of computing power."[99]

The more that tech panic discourse permeates the media, the more pressure politicians feel to act. Of course, such public pressure is not a bad thing in and of itself. Politicians and policymakers should listen and respond to those who have elected them. Thus, it becomes the responsibility of more sober-minded experts to ensure that their voices are heard.

As politicians react, however, they will react with regulatory proposals that aim to curb the harm perceived as the most prominent. Basing public policies on peak fears has driven some of the worst laws and measures in United States history. For instance, the fear of Japanese people living in the US during World War II led to Japanese internment camps, and the fear of terrorism in the

98 Adam Thierer, "Here Come the Code Cops: Senate Hearing Opens Door to FDA for Algorithms & AI Occupational Licensing," *Medium*, May 16, 2023, https://medium.com/@AdamThierer/here-come-the-code-cops-senate-hearing-opens-door-to-fda-for-algorithms-ai-occupational-65b16d8f587d.

99 Billy Perrigo, "Exclusive: U.S. Must Move 'Decisively' to Avert 'Extinction-Level' Threat From AI, Government-Commissioned Report Says," *Time*, March 11, 2024, https://time.com/6898967/ai-extinction-national-security-risks-report/.

immediate aftermath of 9/11 led to domestic spying programs such as the Patriot Act. A precautionary approach has costs of its own, including forgone innovation and other curtailments of commerce, creativity, or even free speech.[100] The small size of the European digital technology industry serves as a prime example of the results of such precaution when it is translated into widespread restrictions on innovative activities.[101]

Lately, academics and organizations focused on x-risk have been escalating their calls for extreme political and regulatory interventions, and some of their ideas now serve as the baseline in public policy debates about artificial intelligence. The work of many of these individuals and groups can be traced to proposals set forth by Nick Bostrom, director of the Future of Humanity Institute at the University of Oxford. Bostrom has done influential writing and speaking on existential risk (or what he calls "superintelligence") and potential global regulatory responses to it. He has outlined a variety of specific regulatory options for addressing existential concerns, most notably in his widely cited essay developing what he refers to as his "vulnerable world hypothesis."[102]

Bostrom's approach to x-risk basically suggests that it is worth pursuing one sort of existential risk (global authoritarian control of science, innovation, and individuals) to address what he regards as a far greater existential risk (the development of

100 Adam Thierer, "Getting AI Innovation Culture Right" (R Street Policy Study No. 281, R Street Institute, Washington, DC, March 2023), https://www.rstreet.org/research/getting-ai-innovation-culture-right.

101 Adam Thierer, "Why the Future of AI Will Not Be Invented in Europe," *Technology Liberation Front*, August 1, 2022, https://techliberation.com/2022/08/01/why-the-future-of-ai-will-not-be-invented-in-europe.

102 Nick Bostrom, "The Vulnerable World Hypothesis," *Global Policy* 10, no. 4 (November 2019): 455–76.

dangerous autonomous systems). During a 2019 TED talk, Bostrom said that ubiquitous mass surveillance might need to be accomplished through global government solutions that could possibly include a "freedom tag" or some sort of "necklace with multi-dimensional cameras" that allows real-time monitoring of citizens to ensure they are not engaged in risky activities.[103] He admitted that there are "huge problems and risks" associated with the idea of mass surveillance and global governance, but suggested that "we seemed to be doomed anyway" and that extreme solutions are acceptable in that light.[104]

Most other AI x-risk theorists do not go quite as far as Bostrom, but many of them also call for fairly sweeping regulatory solutions, some of which entail some sort of global government-imposed regulations. While these proposed solutions are often highly aspirational and lack details, there have been calls for governments to engage in chip-level surveillance using some sort of tracking technology embedded in semiconductors that power large-scale computing systems.[105] This would require some government or organization to track chip distribution and usage in real time across the globe in order to determine how chips are being used and ensure compliance with whatever restrictions on use

103 Nick Bostrom, "How Civilization Could Destroy Itself—and 4 Ways We Could Prevent It," TED, April 2019, https://www.ted.com/talks/nick_bostrom_how_civilization_could_destroy_itself_and_4_ways_we_could_prevent_it?language=en.

104 Bostrom, "How Civilization Could Destroy Itself."

105 Yonadav Shavit, "What Does It Take to Catch a Chinchilla? Verifying Rules on Large-Scale Neural Network Training via Compute Monitoring," *arXiv*, May 30, 2023, https://arxiv.org/abs/2303.11341.

are devised.[106] Extensive software export controls would probably accompany such regulations.[107]

Other academics and organizations have proposed mandating "know your customer" regulations or other supply-chain regulations that would require companies to report their customers (or their customers' activities) to government officials.[108] Another type of proposed regulation would impose hard caps of the aggregate amount of computing power of AI models..[109] Such approvals would be obtained from a new licensing regime that would place limits on who could develop high-powered computing systems.[110]

When Dan Hendrycks, the initiator of the second x-risk open letter, was asked about this letter, he explained that it may take a warning shot—a near disaster—to get the attention of a broad audience. To help the world understand the danger as he does. Hendrycks hopes for a multinational regulation that would include China: "We might be able to jointly agree to slow down." He imagines something similar to the European Organization for Nuclear Research, or CERN. According to Hendrycks, if the private

106 Lennart Heim, "Video and Transcript of Presentation on Introduction to Compute Governance," *Heim.xyz*, May 17, 2023, https://blog.heim.xyz/presentation-on-introduction-to-compute-governance/.

107 Luke Muehlhauser, "12 Tentative Ideas for US AI Policy," Open Philanthropy, April 17, 2023, https://www.openphilanthropy.org/research/12-tentative-ideas-for-us-ai-policy.

108 Muehlhauser, "12 Tentative Ideas for US AI Policy"; "Governing AI: A Blueprint for the Future," Microsoft, May 25, 2023, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw.

109 Kelsey Piper, "A.I. Is About to Get Much Weirder. Here's What to Watch For," *The New York Times*, March 21, 2023, https://www.nytimes.com/2023/03/21/opinion/ezra-klein-podcast-kelsey-piper.html.

110 Ardi Janjeva et al., "Strengthening Resilience to AI Risk: A Guide for UK Policymakers" (Briefing Paper, Centre for Emerging Technology and Security, Alan Turing Institute, August 2023), 36, https://cetas.turing.ac.uk/publications/strengthening-resilience-ai-risk; Markus Anderljung et al., "Frontier AI Regulation: Managing Risks to Public Safety," *arXiv*, July 6, 2023, https://arxiv.org/abs/2307.03718.

sector remains in the lead, the governments of the United States, England, and China could build an "off-switch."[111] (At the same time, Hendrycks is an adviser to Elon Musk's new AI start-up, xAI.)[112]

Moreover, some analysts suggest that "there's only one way to control AI: Nationalization."[113] In their view, governments should consider nationalizing supercomputing facilities, perhaps through a "Manhattan Project for AI Safety," which would be a government-controlled lab that has exclusive authority to coordinate and conduct "high-risk R&D."[114] Others have floated the idea of accomplishing this at a global scale through a new super-regulator that would "remove research on powerful, autonomous AI systems away from private firms and into a highly-secure facility with multinational backing and supervision."[115]

The most radical proposal along these lines comes from Ian Hogarth, who has called for "governments to take control by regulating access to frontier hardware" to limit what he calls "God-like AI."[116] He advocates that such systems be contained on a hypothetical "island," where "experts trying to build God-like [artificial general intelligence] systems do so in a highly secure facility:

111 David Scharfenberg, "Dan Hendrycks Wants to Save Us from an AI Catastrophe. He's Not Sure He'll Succeed," *The Boston Globe*, July 6, 2023, https://www.bostonglobe.com/2023/07/06/opinion/ai-safety-human-extinction-dan-hendrycks-cais/.

112 Sharon Goldman, "Doomer AI Advisor Joins Musk's xAI, the 4th Top Research Lab Focused on AI Apocalypse," *VentureBeat*, July 24, 2023, https://venturebeat.com/ai/doomer-advisor-joins-musks-xai-the-4th-top-research-lab-focused-on-ai-apocalypse/.

113 Charles Jennings, "There's Only One Way to Control AI: Nationalization," *Politico*, August 20, 2023, https://www.politico.com/news/magazine/2023/08/20/its-time-to-nationalize-ai-00111862.

114 Samuel Hammond, "A Manhattan Project for AI Safety," *Second Best*, May 15, 2023, https://www.secondbest.ca/p/a-manhattan-project-for-ai-safety.

115 Andrea Miotti, "Priorities for the UK Foundation Models Taskforce," *Ars Longa, Vita Brevis*, July 23, 2023, https://andreamiotti.substack.com/p/uk-taskforce-priorities.

116 Hogarth, "We Must Slow Down the Race to God-Like AI."

an air-gapped enclosure with the best security humans can build. All other attempts to build God-like AI would become illegal; only when such AI were provably safe could they be commercialized 'off island.'"[117] As mentioned, his proposal gained a lot of attention.

To be clear, under this and other nationalization schemes, private commercial development of advanced supercomputing systems and models would be illegal. Incredibly, none of the authors of these proposals have anything to say about how they plan to convince China or other nations to abandon all their supercomputing facilities and research. Such cooperation would be required for "island" schemes to have any serious global limiting effect.

Still other analysts speak of the need for a "public option for superintelligence" that would have governments exert far greater control over large-scale generative AI systems, perhaps by creating their own publicly funded systems or models.[118]

While most governments have yet to act on these calls, some lawmakers are threatening far-reaching controls on computation and algorithmic innovations for risks more mundane than "superintelligence." For example, Italy banned ChatGPT for a month in April 2023 over privacy concerns before finally allowing OpenAI to restore service to the country.[119] In the US, greatly expanded legal liability is being proposed as a solution to hypothetical harms that have not yet developed. Sen. Josh Hawley for reasons

117 Hogarth, "We Must Slow Down the Race to God-Like AI."

118 Jack Clark, "What Should the UK's £100 Million Foundation Model Taskforce Do?," *Import AI*, July 5, 2023, https://jack-clark.net/2023/07/05/what-should-the-uks-100-million-foundation-model-taskforce-do/; Bruce Schneier and Nathan E. Sanders, "Build AI by the People, for the People," *Foreign Policy*, June 12, 2023, https://foreignpolicy.com/2023/06/12/ai-regulation-technology-us-china-eu-governance.

119 Adi Robertson, "ChatGPT Returns to Italy after Ban," *Verge*, April 28, 2023, https://www.theverge.com/2023/4/28/23702883/chatgpt-italy-ban-lifted-gpdp-data-protection-age-verification.

of "privacy . . . the harms of unchecked AI development, insulate kids from harmful impacts, and keep[ing] this valuable technology out of the hands of our adversaries" set forth objectives for AI legislation that expanded lawsuits for AI models, and he also proposed a new federal regulatory licensing regime for generative AI.[120] Along with Sen. Richard Blumenthal, Senator Hawley also sent a letter to Meta in June 2023, citing "spam, fraud, malware, privacy violations, harassment, and other wrongdoing and harms" and warning the company about how it released its open-sourced "LLaMA" model. The letter even suggested that closed-source models were preferable.[121]

Meanwhile, in the UK, Prime Minister Rishi Sunak commented on the widely circulated second open letter, saying, "The government is looking very carefully at this."[122] To prove as much, in June the British government appointed Ian Hogarth, author of the "AI island" proposal discussed earlier, to lead its new AI Foundation Model Taskforce.[123]

The danger with extreme political solutions to hypothetical AI risks is not only that these reactions could derail many beneficial forms of innovation (especially open-source AI innovation),[124] but—more importantly—that they require profoundly dangerous trade-offs

120 Josh Hawley, "Hawley Announces Guiding Principles for Future AI Legislation," Josh Hawley's Senate website, June 7, 2023, https://www.hawley.senate.gov/hawley-announces-guiding-principles-future-ai-legislation.

121 Richard Blumenthal and Josh Hawley to Mark Zuckerberg, June 6, 2023, https://www.blumenthal.senate.gov/imo/media/doc/06062023metallamamodelleakletter.pdf.

122 Quoted in Laurie Clarke and Annabelle Dickson, "Sunak and Biden to Discuss AI after 'Extinction Risk' Warning," *Politico*, May 31, 2023, https://www.politico.eu/article/sunak-and-biden-to-discuss-ai-after-extinction-risk-warning.

123 UK government, "Tech Entrepreneur Ian Hogarth to Lead UK's AI Foundation Model Taskforce," press release, June 18, 2023, https://www.gov.uk/government/news/tech-entrepreneur-ian-hogarth-to-lead-uks-ai-foundation-model-taskforce.

124 Adam Thierer, "Will AI Policy Became [*sic*] a War on Open Source Following Meta's Launch of LLaMA 2?," *Medium*, July 18, 2023, https://medium.com/@AdamThierer/will-ai-policy-became-a-war-on-open-source following-metas-launch-of-llama-2-b713a3dc360d.

in the realms of human rights and global stability. Bostrom and many other advocates of global regulatory interventions to address what they perceive as serious risks should consider what we can learn from the past and especially from previous calls for sweeping global controls to address new innovations.

At the outset of the Cold War, for example, the real danger of nuclear escalation among superpowers led some well-meaning intellectuals to call for extreme steps to address the existential risk associated with global thermonuclear conflict. In 1951, the eminent philosopher Bertrand Russell predicted "the end of human life, perhaps of all life on our planet," before the end of the century unless the world unified under "a single government, possessing a monopoly of all the major weapons of war."[125] Fortunately, Russell's recommendations were not heeded. Instead, the risks from nuclear weapons are managed in a multistakeholder, voluntary, and (mostly) peaceful manner. Despite the vast difference between nuclear weapons and AI, many AI x-risk theorists today similarly imagine that only sweeping global governance solutions can save humanity from near-certain catastrophe. If the risks from weapons of mass destruction have been managed successfully (to date), then the likelihood of successful management is much greater for a nonweapon technology like AI. To be clear, this paper is not drawing a comparison between the risk of nuclear weapons and the risk of AI, because such a comparison would be inaccurate and unhelpful.[126] The point is that past powerful technologies have also prompted calls for draconian regulations.

125 Bertrand Russell, "The Future of Man," *The Atlantic*, March 1951, https://www.theatlantic.com/magazine/archive/1951/03/the-future-of-man/305193.

126 Hodan Omaar, "No We Aren't in an Oppenheimer Moment for AI," Center for Data Innovation, July 28, 2023, *https://datainnovation.org/2023/07/no-we-arent-in-an-oppenheimer-moment-for-ai/*.

What scholars then and now have failed to address fully are the remarkable and quite conceivable dangers associated with their proposals. "Global totalitarianism is its own existential risk," notes researcher Maxwell Tabarrok regarding Bostrom's approach.[127] Indeed, the threat to liberties and lives from totalitarian government was a sad historical legacy of the past century. An effort to create a single global government authority to oversee all algorithmic risks could lead to serious conflict among nations vying for that power, as well as to mass disobedience by nation-states, companies, organizations, and individuals who want no part of such a scheme.

Extreme rhetoric has undermined life-enriching innovation before. For example, fears of nuclear weapons have contributed to generalized fears of nuclear power and radiation. This has resulted in the multi-decade stifling of cheap, abundant energy derived from nuclear fission. These fears still drive misguided efforts to take existing nuclear plants offline, which actually costs more lives than it saves.[128]

AI has been referred to as a "general purpose technology," which means that it is a technology applicable in many use cases. Electricity is also a general-purpose technology. When electric power was first deployed, there was a massive panic about electricity.[129] Had policymakers acted on those fears and restricted the development and use of electricity, their actions would have had widespread

127 Maxwell Tabarrok, "Enlightenment Values in a Vulnerable World," *Effective Altruism Forum*, July 18, 2022, https://forum.effectivealtruism.org/posts/A4fMkKhBxio83NtBL/enlightenment-values-in-a-vulnerable-world.

128 Adam Thierer, "How Many Lives Are Lost Due to the Precautionary Principle?," *Discourse*, October 31, 2019, https://www.mercatus.org/economic-insights/expert-commentary/how-many-lives-are-lost-due-precautionary-principle.

129 Joseph P. Sullivan, "Fearing Electricity: Overhead Wire Panic in New York City," *IEEE Technology and Society* 14 (Fall 1995): 8–16.

and quite deleterious consequences for society. Instead, people quickly came to understand and experience the benefits of electric power. If AI panic drives extreme political responses to AI, these responses could also have negative ramifications.

Finally, practically speaking, any attempt to create global government solutions to mitigate AI risks must contend with the fact that most experts cannot even agree on how to define artificial intelligence. Nor is there any clear consensus about what are the most serious algorithmic dangers that should be addressed through global accords or even domestic regulations. There is a major divide currently between those in the "AI ethics" camp and those in the "AI safety" camp, and it has led to heated arguments about what issues deserve the most attention.[130] Elevating AI policy battles to a global scale would multiply the range of issues in play and of actors who want some control over decision-making, creating the potential for even more conflict.

Around the globe, media hype influences politicians and their proposals to regulate AI. The rhetoric is not without consequence, and the results will continue to unfold. It is up to voters and the policymakers themselves to avoid the hype and focus on the issues that are actively harming consumers. This approach will lead the creation of more sober and innovation-friendly policy while allowing governments at all levels to step in and correct harms.

130 Emily M. Bender, "Talking about a 'Schism' Is Ahistorical," *Medium*, July 5, 2023, https://medium.com/@emilymenonbender/talking-about-a-schism-is-ahistorical-3c454a77220f.

Effect on the Public

The public also responds to negative AI hype, dire predictions, and extreme proposals.

Existential risk, once a niche discussion, has gained popular prominence. For example, in May 2023, a *Reuters* poll revealed that 61 percent of Americans think that AI poses an existential risk to humanity.[131] A May 2023 Quinnipiac University poll showed that "a majority of Americans (54 percent) think artificial intelligence poses a danger to humanity, while 31 percent think it will benefit humanity."[132] According to an April 2023 survey by Morning Consult, two out of three (61 percent) adults in the United States now perceive AI tools to be an existential threat to humanity.[133]

Before ChatGPT, AI was at the bottom of Americans' list of risk concerns. A survey by the Centre for the Governance of AI measured the public's perception of the global risk of AI within the context of other global risks by asking respondents to respond to questions about cyberattacks, terrorist attacks, global recession, and the spread of infectious diseases. The centre defined "global risk" as "an uncertain event or condition that, if it happens, could cause significant negative impact for at least 10 percent of the

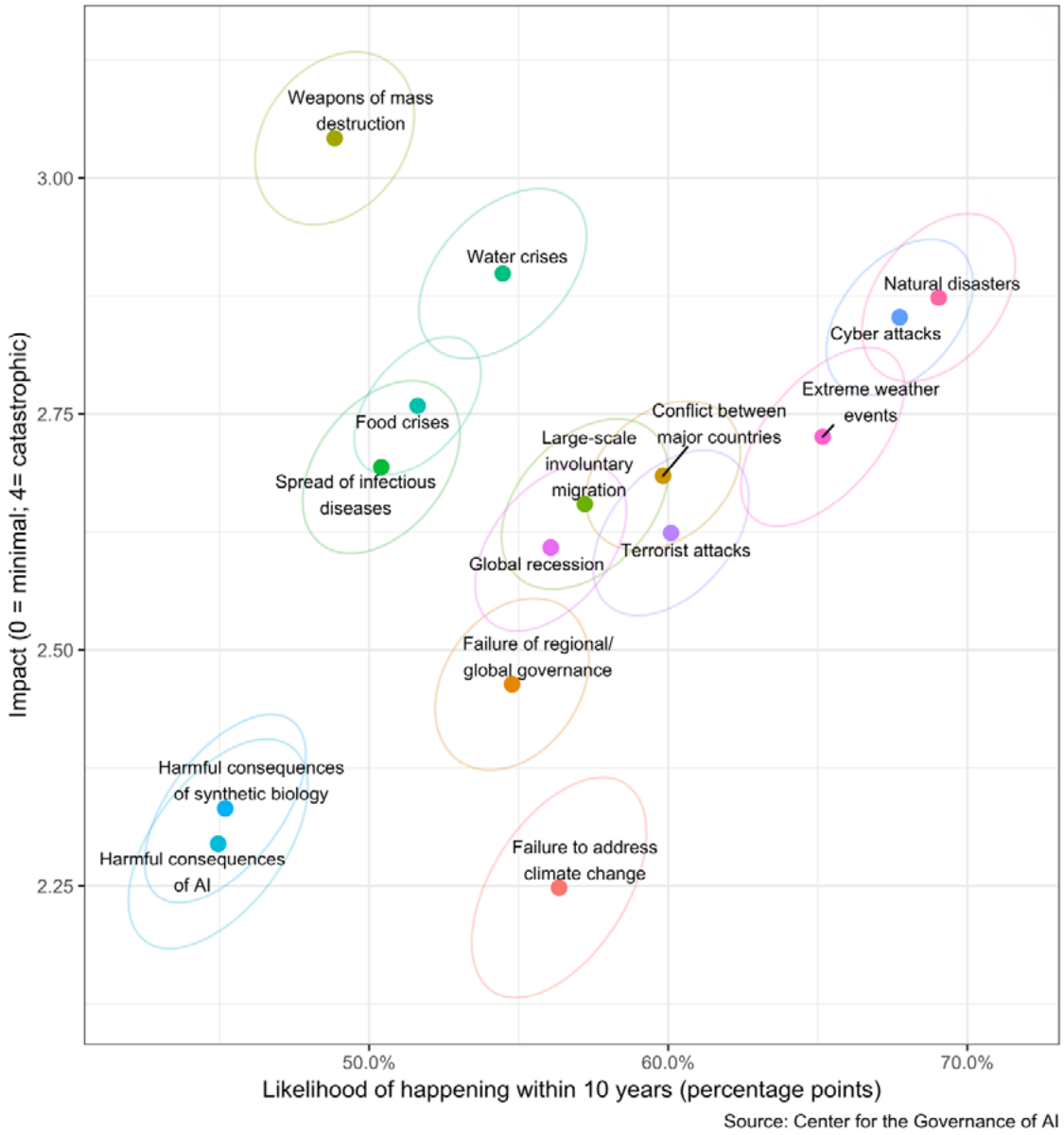
131 Anna Tong, "AI threatens humanity's future, 61% of Americans say: Reuters/ Ipsos poll," *Reuters*, May 17, 2023, https://www.reuters.com/technology/ai-threatens-humanitys-future-61-americans-say-reutersipsos-2023-05-17/

132 Quinnipiac University. "Trump Doubles Lead Over DeSantis in 2024 GOP Primary Race, Quinnipiac University National Poll Finds; 65% Of Voters Think Biden Is Too Old For Second Term." May 24, 2023. https://poll.qu.edu/poll-release?releaseid=3872

133 Dellinger, AJ. "Most of the Public Believes Artificial Intelligence Tools Can Achieve Singularity and Pose a Threat to Humanity." Morning Consult. April 12, 2023. https://pro.morningconsult.com/instant-intel/generative-ai-singularity

world's population."[134] The findings from the survey are reported in the scatterplot shown in figure 1.

FIGURE 1 | AI In the Context of Other Global Risk



Source: Baobao Zhang and Allan Dafoe, "Artificial Intelligence: American Attitudes and Trends," Center for the Governance of AI, Future of Humanity Institute, University of Oxford May 1, 2023, 3.

134 Baobao Zhang and Allan Dafoe, "Artificial Intelligence: American Attitudes and Trends." Center for the Governance of AI, Future of Humanity Institute, University of Oxford. (January 2019). https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/us_public_opinion_report_jan_2019.pdf.

The data reported in figure 1 clearly show that respondents ranked AI risk lowest overall in 2019 along both axes. On the horizontal "Likelihood" axis, the "harmful consequences of AI" was the least probable event to occur over the next 10 years. Along the vertical "Impact" axis it ranked toward the bottom of potential impacts, above only "failure to address climate change."

A survey conducted by researchers from Monmouth University is perhaps the most insightful. Monmouth conducted this survey in 2015 and again in 2023, asking respondents various questions about their opinions regarding the increased adoption of AI systems and the potential consequences of this increase. In both years, the survey asked, "How worried are you that machines with artificial intelligence could eventually pose a threat to the existence of the human race?" The 2015 poll found that 44 percent of respondents were either "very worried" or "somewhat worried," while 56 percent were either "not at all worried" or "not too worried." The 2023 poll saw the "very worried" and "somewhat worried" categories jump to 56 percent and the "not at all worried" and "not too worried" categories fall to 44 percent.[135] Similarly, a Pew Research Center survey from August 2023 found that 52 percent of Americans say they feel more concerned than excited about the increased use of AI. This was up 14 percent since December 2022, when 38 percent expressed this concern.[136]

Evidently, optimism surrounding AI technologies used to be higher. We should look at the fruits of AI technological development

135 Monmouth University Polling Institute. "Artificial Intelligence Use Prompts Concerns." February 15, 2023. https://www.monmouth.edu/polling-institute/reports/monmouthpoll_US_021523/.

136 Alec Tyson and Emma Kikuchi. "Growing public concern about the role of artificial intelligence in daily life." Pew Research. August 28, 2023. https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/.

with awe, but allowing the doomsday conversation to dominate the public consciousness may lead to more negative externalities than the real probability of an AI overlord.

Recommendations

Current media coverage is amplifying the potential existential risk from AI. This is unsurprising, because the media thrives on fear-based content. However, we can expect that the doomsaying will *not* stay this dominant. Panic cycles, as their name implies, are circular. At some point, the hysteria calms down (see figure 2).[137] Here are some suggestions for ways to reach that point:

For the media:

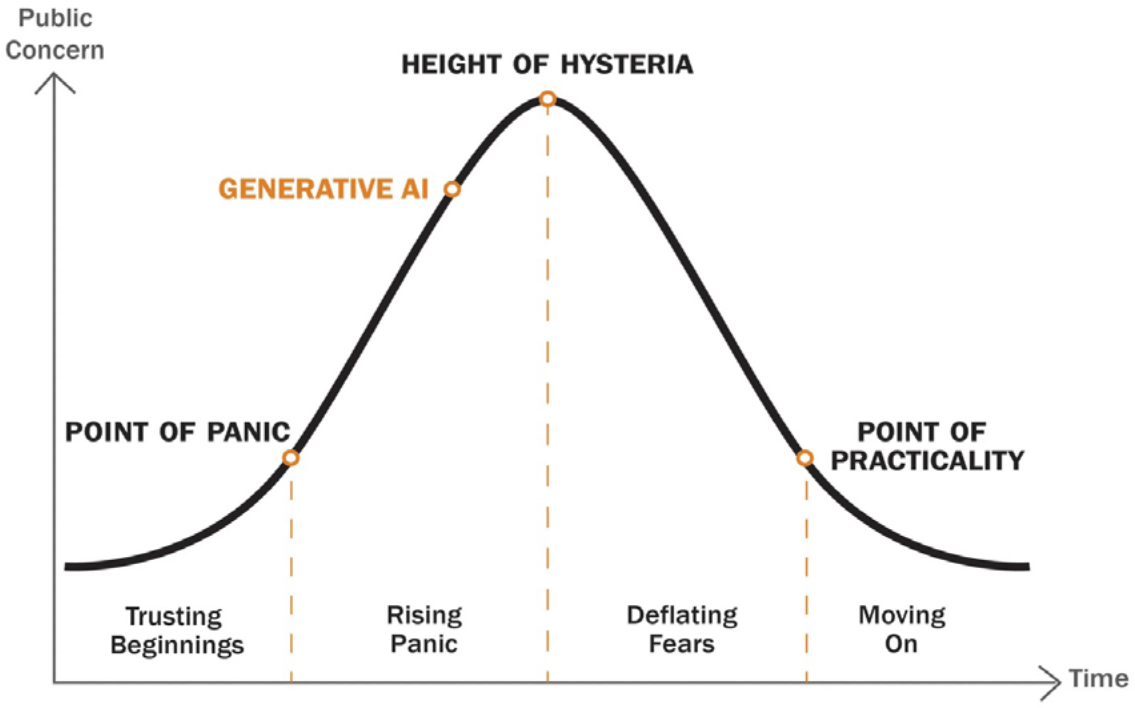
→ The media needs to stop spreading unrealistic expectations (both good and bad). The focus should be on how AI systems actually work (and don't work). When we discuss what AI is, we also need to discuss what it isn't.

→ Media attention should not be paid to the fringes of the debate. The focus should return to the actual challenges and the guardrails they require.

→ There are plenty of AI researchers who would love to inform the public in a nuanced way. It's time to highlight more diverse voices that can offer different perspectives.

137 Patrick Grady and Daniel Castro, "Tech Panics, Generative AI, and the Need for Regulatory Caution," Center for Data Innovation, May 1, 2023, 3, https://datainnovation.org/2023/05/tech-panics-generative-ai-and-regulatory-caution/.

FIGURE 2 | The AI Panic Cycle: Fears Increase, Peak, Then Decline over Time as the Public Becomes Familiar with the Technology and Its Benefits



Source: Patrick Grady and Daniel Castro, "Tech Panics, Generative AI, and the Need for Regulatory Caution," Center for Data Innovation, May 1, 2023, 3.

For media audiences:

→ Whenever people make sweeping predictions with absolute certainty in a state of uncertainty, it is important to raise questions about what motivates such extreme forecasts.

→ We need to keep reminding ourselves that the promoters of hype and criti-hype[138] have much to gain from spreading the impression that AI is much more powerful than it actually is. Rather than getting caught up in these hype cycles, we should be skeptical and evaluate in a more nuanced way how AI affects our daily lives.

138 Lee Vinsel, "You're Doing It Wrong: Notes on Criticism and Technology Hype," *Medium*, February 1, 2021, https://medium.com/p/18b08b4307e5.

→ We need to look at the complex reality and see humans at the helm, not machines. It's humans making decisions about the design, training, and applications. Many social forces are at play here: researchers, policymakers, industry leaders, journalists, and users all have a hand in shaping this technology.

For policymakers:

→ First, policymakers should be aware of current technopanics and respond accordingly. Patrick Grady and Daniel Castro urge, "It would behoove policymakers to recognize when they are in the midst of a tech panic and use caution when digesting hypothetical or exaggerated concerns about generative AI that crowd out discussion of more immediate and valid ones."

→ Second, policymakers should base their policy recommendations and decisions on actual harms, not hypothetical ones. As noted earlier, panic-fueled public policy decisions have a history of negative effects.[139]

→ Third, policymakers should use their public platforms to educate the public about the actual technology in play. Policymakers are uniquely situated in that they are able to solicit and hear from a wide array of experts. Although they face a range of incentives that might run counter to this recommendation, they have a responsibility to provide a sober analysis. Sen. Chuck Schumer and Sen. Bill Cassidy are excellent examples of

139 Patrick Grady and Daniel Castro, "Tech Panics, Generative AI, and the Need for Regulatory Caution," *Center for Data Innovation*, May 1, 2023, https://datainnovation.org/2023/05/tech-panics-generative-ai-and-regulatory-caution/.

prominent politicians using their platform to slowly assess the issue and educate their colleagues and the public.[140] [141]

Conclusion

Extreme rhetoric about AI is ubiquitous and has a real influence on politicians and public opinion. The danger is that this rhetoric will result in policy decisions that come at the cost of potentially lifesaving technologies. When a technology is as important to the economy as AI, the incentives of the x-risk institutions and the effects of their rhetoric are worth further examination. Moving forward, the solutions to dealing with this media hype should be multifaceted and should involve the whole of civil society.

140 United States Senate. "Majority Leader Schumer Floor Remarks On The Senate's First AI Insight Forum To Take Place Next Week." September 7, 2023. https://www.democrats.senate.gov/newsroom/press-releases/majority-leader-schumer-floor-remarks-on-the-senates-first-ai-insight-forum-to-take-place-next-week.

141 U.S. Senate Committee On Health, Education, Labor, and Pensions. "Ranking Member Cassidy Releases White Paper on Artificial Intelligence." September 6, 2023. https://www.help.senate.gov/ranking/newsroom/press/ranking-member-cassidy-releases-white-paper-on-artificial-intelligence.