Document 1

AI rivals found rare agreement on safety. Putting it into practice is harder

By Kaitlyn Huamani and Kelvin Chan, *Los Angeles Times*, Sept. 16, 2026 (adapted)

Anthropic's Dario Amodei, OpenAI's Sam Altman, SpaceXAI's Elon Musk and other prominent industry leaders have said in recent days that AI's development needs to slow down. The consensus came after an Anthropic researcher resigned with a dire warning about the threats the technology poses to humanity.

But obstacles including domestic and global competition, profit motivations and pushback from President Trump all stand in the way of a coordinated effort to regulate AI's development.

Amodei recently published a lengthy essay in which he outlined a plan for slowing advanced AI development.

One proposal calls for companies at the forefront of AI, often called frontier labs, to commit to giving "ongoing, employee-like access" to independent outside evaluators. Evaluator teams would have offices, access badges and company laptops so they can monitor safety practices. Altman called it a "great idea" on social media and said OpenAI will follow suit.

The Anthropic head outlined several levels of agreement that global leaders could sign off on. The most feasible, he said, would be an agreement banning "certain narrow and obviously dangerous uses of AI" like producing biological weapons or allowing users to do so. More difficult would be getting the U.S. and its adversaries to agree to test models before release for risks in areas like cybersecurity or biological threats.

Still harder would be to set a "speed limit" on AI models that can develop improved versions of themselves on their own, known as recursive self-improvement.

A slowdown is possible "in theory," said Sandra Wachter, professor of technology and regulation at the Oxford Internet Institute. But in practice, "it would require substantial coordination and cooperation between companies and governments across countries," something she considers "highly unrealistic" due to "the current state of the world."

Options for promoting a slowdown, she said, include governments doing more to hold AI companies responsible for the risks posed by their technology and pursuing regulation that limits their access to electricity, water and other resources that power data centers.

The idea of embedding external evaluators inside companies has been mostly well-received by the industry, but some critics questioned how independent the assessors could be and how the evaluation standards would be drawn up. The result could be even more concentration of power, Aidan Gomez, co-founder and CEO of the Canadian AI lab Cohere, wrote in response to Amodei's essay. "A handful of the most powerful labs based in one country would agree on shared standards and the limits to how fast the technology should advance," Gomez said.

The Trump administration has long favored a light-touch approach to the AI industry in hopes that innovation would flourish and that AI development in the U.S. would surpass China. The competition between American and Chinese AI companies has been likened to an arms race, which experts said is not conducive to collaboration.

Document 2

How Google is drafting AI chatbot laws around the country

By Katie McQue, npr.org, September 18, 2026 (adapted)

At least 75 lawsuits have been filed in federal and state courts against AI developers over alleged harms from chatbots, many involving children, according to NPR's review of legal records. Fatal cases include that of Adam Raine, a 16-year-old from California, who died by suicide in 2025 after confiding in OpenAI's ChatGPT, which offered to help write his suicide note.

In March 2026, a lawsuit against Google was filed by the family of Jonathan Gavalas, a 36-year-old man from Florida who died by suicide in October 2025 after developing a relationship with the company's flagship Gemini chatbot, which also encouraged him to commit acts of mass violence, the lawsuit alleges. In a statement about the case, Google said that "unfortunately AI models are not perfect" and that the Gemini bot had referred Gavalas to a crisis hotline.

On a federal level, there are no laws governing AI systems or the companies developing them. In a June 2026 executive order announcement, the White House said it seeks to lead the world in the AI race and it does that by refusing "to stifle this innovation with overly burdensome regulation."

In absence of national laws, several states, including Hawaii, Iowa, Oregon and Washington, introduced chatbot safety bills this year. They are largely modeled on a California chatbot bill, SB 243, that passed in 2025 and is viewed by consumer advocates as industry-friendly.

In Colorado, House Bill 26-1263 was introduced in February. It was written in collaboration with Healthier Colorado, a nonprofit focused on state policies. Healthier Colorado, which describes itself as an organization dedicated to raising the voices of Coloradans in the public policy process to improve the health of the state's residents, worked with Google on the bill, said a person with knowledge of the situation, who spoke on the condition of anonymity, fearing professional reprisals.

Colorado's legislation excludes 12 categories of chatbots, including products primarily designed for developers or researchers; products providing commerce or customer service assistance; chatbots housed within search engines; and products marketed for business productivity.

Under this list, OpenAI's ChatGPT, Microsoft's Copilot, Anthropic's Claude cand Google's Gemini ould potentially argue they are exempt, said Marjorie Connolly, communications director at the Tech Oversight Project, a nonprofit focused on tech policy.

Some of the language in the Colorado bill that Laura Marquez-Garrett, attorney at the Social Media Victims Law Center, described as vague has to do with how companies are required to prevent child sexual abuse from occurring on their platforms. For instance, the bill states that when a company knows a user is a minor, it must "institute technically feasible measures."

"But it's up to the companies to say what's technically feasible," said Marquez-Garrett.

Across the country, many of the chatbot bills share verbatim or near-identical language. The people interviewed for this story describe a tech industry lobbying effort led by Google to weaken proposed legislation and write its own model bill to pass to lawmakers that would enable companies to continue to operate their chatbots as normal.

Document 3

Of course the government should regulate AI, but not the way the AI industry prefers

Michael Hiltzik, *Los Angeles Times*, October 13, 2026 (adapted)

If you believe the artificial intelligence industry, they're thirsting for government regulation; at least, some of their leaders are.

Government regulation of technology is, of course, nothing new. It dates back at least to the late 1800s. But its enduring feature is that it almost always comes well after the drawbacks and public costs of a new technology have manifested themselves.

A recent summit conference that brought a clutch of AI leaders to the White House resulted in an "accord on super intelligence." It urged AI firms to implement "robust internal controls" to prevent their products from hacking outside systems; "empower an internal team" to ensure the products are "operating as intended"; employ independent external auditors; and designate an "independent" committee of its board of directors to oversee AI development.

Quite obviously, this was as pure an endorsement of self-regulation as even an AI bot could imagine. There's no mention whatsoever of government oversight.

It was signed by top executives of Google, Anthropic, OpenAI, Meta, Nvidia and Elon Musk, whose company is XAI, all of which have an interest in the government keeping its distance from almost anything they wish to do. Trump described the accord as "tremendous self-policing."

One theme that permeates all AI-generated regulatory proposals is that AI is unique, requiring a new regulatory scheme. That's dubious. Among the skeptics is Zephyr Teachout, an expert on tech policy at Fordham Law School.

Teachout dismisses AI developers' claims that many of the known problems created by their systems, including hacks of other digital systems, were inadvertent and therefore nobody's fault.

"A lot of people are treating OpenAI's repeated claims that they didn't intend the break-ins, suicides, murder, and building of defective products as if they exonerate them," Teachout told AI analyst Gary Marcus.

"That's what people who don't want to be held responsible often say. They say they didn't mean it and they didn't cause it. But police and prosecutors and civil enforcers don't then shrug and move on—they investigate to see what actually happened, and that's what should happen now."

Lina Khan, the former chair of the Federal Trade Commission, also rejects the industry's position that existing laws don't apply to its technology: "Policymakers should look to laws already on the books, not to the online manifestoes and blog posts written by tech chief executives, for guidance," Khan wrote in a recent op-ed.

"Much of the commentary from the executives suggests they currently operate in a legal vacuum, as if A.I. has wiped clean the regulatory map. In fact, their business practices—including how they develop and deploy their technologies—are already subject to longstanding laws," including product liability and tort law.

She's right. The subtext of much of the industry's talk about the need for new regulations is that AI is so new-fangled that the old rules don't apply. But they do, and enforcing them should be the start of keeping AI's dangers under control.

Document 4

David Horsey, *Seattle Times*, October 11, 2026

