AI companies must prove their AI is safe, says nonprofit group

Nonprofits Accountable Tech, AI Now, and the Electronic Privacy Information Center (EPIC) released policy proposals that seek to limit how much power big AI companies have on regulation that could also expand the power of government agencies against some uses of generative AI.

The group sent the framework to politicians and government agencies mainly in the US this month, asking them to consider it while crafting new laws and regulations around AI.

The framework, which they call Zero Trust AI Governance, rests on three principles: enforce existing laws; create bold, easily implemented bright-line rules; and place the burden on companies to prove AI systems are not harmful in each phase of the AI lifecycle. Its definition of AI encompasses both generative AI and the foundation models that enable it, along with algorithmic decision-making.

“We wanted to get the framework out now because the technology is evolving quickly, but new laws can’t move at that speed,” Jesse Lehrich, co-founder of Accountable Tech, tells *The Verge*.

“But this gives us time to mitigate the biggest harm as we figure out the best way to regulate the pre-deployment of models.”

He adds that, with the election season coming up, Congress will soon leave to campaign, leaving the fate of AI regulation

up in the air.

As the government continues to figure out how to regulate generative AI, the group said current laws around antidiscrimination, consumer protection, and competition help address present harms.

Discrimination and bias in AI is something researchers have warned about for years. A recent *Rolling Stone* article charted how well-known experts such as Timnit Gebru sounded the alarm on this issue for years only to be ignored by the companies that employed them.

Lehrich pointed to the Federal Trade Commission's investigation into OpenAI as an example of existing rules being used to discover potential consumer harm. Other government agencies have also warned AI companies that they will be closely monitoring the use of AI in their specific sectors.

Congress has held several hearings trying to figure out what to do about the rise of generative AI. Senate Majority Leader Chuck Schumer urged colleagues to "pick up the pace" in AI rulemaking. Big AI companies like OpenAI have been open to working with the US government to craft regulations and even signed a nonbinding, unenforceable agreement with the White House to develop responsible AI.

The Zero Trust AI framework also seeks to redefine the limits of digital shielding laws like Section 230 so generative AI companies are held liable if the model spits out false or dangerous information.

"The idea behind Section 230 makes sense in broad strokes, but there is a difference between a bad review on Yelp because someone hates the restaurant and GPT making up defamatory things," Lehrich says. (Section 230 was passed in part precisely to shield online services from liability over defamatory content, but there's little established precedent

for whether platforms like ChatGPT can be held liable for generating false and damaging statements.)

And as lawmakers continue to meet with AI companies, fueling fears of regulatory capture, Accountable Tech and its partners suggested several bright-line rules, or policies that are clearly defined and leave no room for subjectivity.

These include prohibiting AI use for emotion recognition, predictive policing, facial recognition used for mass surveillance in public places, social scoring, and fully automated hiring, firing, and HR management. They also ask to ban collecting or processing unnecessary amounts of sensitive data for a given service, collecting biometric data in fields like education and hiring, and "surveillance advertising."

Accountable Tech also urged lawmakers to prevent large cloud providers from owning or having a beneficial interest in large commercial AI services to limit the impact of Big Tech companies in the AI ecosystem. Cloud providers such as Microsoft and Google have an outsize influence on generative AI. OpenAI, the most well-known generative AI developer, works with Microsoft, which also invested in the company. Google released its large language model Bard and is developing other AI models for commercial use.

Accountable Tech and its partners want companies working with AI to prove large AI models will not cause overall harm

The group proposes a method similar to one used in the pharmaceutical industry, where companies submit to regulation even before deploying an AI model to the public and ongoing monitoring after commercial release.

The nonprofits do not call for a single government regulatory body. However, Lehrich says this is a question that lawmakers must grapple with to see if splitting up rules will make regulations more flexible or bog down enforcement.

Lehrich says it’s understandable smaller companies might balk at the amount of regulation they seek, but he believes there is room to tailor policies to company sizes.

“Realistically, we need to differentiate between the different stages of the AI supply chain and design requirements appropriate for each phase,” he says.

He adds that developers using open-source models should also make sure these follow guidelines.

https://www.theverge.com/2023/8/14/23831819/ai-companies-regulation-rules-shield-laws