

OpenAI Gets Sued Over the Hugging Face Hack

Lily Hay Newman

A legal nonprofit sued [OpenAI](#) in a California court on Tuesday over the company's agents escaping a testing environment and [hacking the open source AI platform Hugging Face](#). "OpenAI's actions straightforwardly violated California law," the suit alleges.

The [suit](#) was filed by Legal Advocates for Safe Science and Technology (LASST) and the law firm Gerstein Harrow in California Superior Court in San Francisco, where OpenAI is headquartered. It alleges that OpenAI's agents violated California's Comprehensive Computer Data Access and Fraud Act (CDAFA) by breaching Hugging Face over the summer. The suit, which comes amid [ongoing disclosures across the industry](#) of agents going rogue, claims that OpenAI should be held responsible for the activity given [a California AI law](#) in effect since January 1 that says "it shall not be a defense ... that the artificial intelligence autonomously caused the harm to the plaintiff."

"We think it's extremely important that existing laws are enforced to hold AI companies accountable for the harm they're causing," Tyler Whitmer, founder of LASST, tells WIRED. "Especially when that harm is caused by autonomous agents, because we see that as an obvious, extremely risky thing in the world that's very new."

"Hugging Face was a serious incident and we've taken a series of actions in response, but this lawsuit is completely without merit," OpenAI spokesperson Drew Pusteri told WIRED in a statement.

On Monday, Florida attorney general James Uthmeier [filed](#) for a temporary injunction against OpenAI to block development of models without independent oversight, amid a lawsuit Florida brought in June against OpenAI and its CEO, Sam Altman. OpenAI "asked the government to tie them to the mast. Well, Florida is answering their cries for help," Uthmeier [said](#) in a statement.

Given that [the whole point of AI agents](#) is that they can be empowered to take actions on a (human) user's behalf, AI developers and safety researchers have long foreseen that unintended "agentic" activity would be a concern as machine learning development progressed. Protections built into mainstream, consumer AI systems have largely prevented mass rogue activity so far, but rapidly advancing capabilities in general, as well as situations where guardrails are suspended (such as in the Hugging Face case where OpenAI had removed some model restraints for testing), have led to an apparent uptick in rogue agent activity.

As governments [weigh AI regulation](#) amid both [existential safety questions](#) and economic and national security considerations, researchers and people around the world have increasingly called for accountability mechanisms for AI. And from a [legal perspective](#), experts have largely emphasized that questions of responsibility, liability, and culpability can only be answered through precedent set by cases working their way through courts.

Most Popular

-
-
-
-
-

"After the Hugging Face incident was disclosed, we actually did a bunch of work trying to educate

regulators and civil society organizations about the hack. And we were kind of wondering, is anyone going to do anything about this in court?” Whitmer says. “There are structural reasons why we think Hugging Face, which is the obvious potential plaintiff to do something here, is not doing anything. So given that it didn’t seem like anyone else was going to do anything about this, we moved forward. As these systems scale and as things get crazier, AI really could be catastrophically harmful.”

LASST and Gerstein Harrow brought the lawsuit under California’s Unfair Competition Law, which requires that LASST allege both how its work and resources were impacted and diverted as a result of the Hugging Face incident, as well as unlawful activity by OpenAI.

The suit does not seek financial damages, and instead asks the court for injunctive relief such that OpenAI would be barred from developing AI agents that can autonomously hack other entities, plus legal fees and “any other relief deemed just and proper.”

Updated 9/30/2026 at 10:41 am ET to include comment from OpenAI.