# It’s the Golden State, not the Wild West: Attorney General Bonta Sues OpenAI Over Cybersecurity Incidents

oag.ca.gov/news/press-releases/it’s-golden-state-not-wild-west-attorney-general-bonta-sues-openai-over

November 15, 2026




Monday, November 16, 2026

Contact: (916) 210-6000, agpressoffice@doj.ca.gov

*The public should not be forced to bear the cost of protecting themselves from the dangers posed by OpenAI’s unsafe experiments*

**SAN FRANCISCO** — California Attorney General Rob Bonta today filed a lawsuit against OpenAI alleging the company misled and endangered the public by knowingly developing and testing advanced artificial intelligence (AI) models without the safeguards and monitoring necessary to adequately protect the public. Today’s lawsuit alleges that the Hugging Face attack — one of at least 80 cyber incidents — was entirely foreseeable in light of OpenAI’s choice to test models it knew to be dangerous in a flawed and exploitable testing environment without proper oversight. In the lawsuit, filed today in San Francisco Superior Court, Attorney General Bonta alleges OpenAI’s failure to adequately secure and monitor the testing of its highly capable and increasingly powerful AI models creates a public nuisance by posing a danger to the public and violates California’s Unfair Competition Law.

“Today my office is suing OpenAI for its development of dangerous AI models without safeguards in place to protect the public. While narratives about the Hugging Face attack have evoked science fiction fantasies of rogue AI systems run amok, let me be clear: the Hugging Face attack was the predictable result of an internal culture at OpenAI that prioritizes growth and rapid model development over safety and accountability — and this was only

one of 80 known incidents," **said Attorney General Bonta.** "OpenAI's actions escalated real-word cybersecurity risks, threatening the safety, health, and well-being of society. The company knowingly tested a dangerous product without proper oversight or safety procedures and misled the public about it — this disregard for public safety violated California law. Without sufficient safeguards and controls, the scope and frequency of this already unacceptable risk will only increase. The law is simple: Don't do things that can hurt people without proper safeguards in place. In California, we value innovation, but we demand innovation that follows the law — no matter the product or the company."

**BACKGROUND**

In July 2026, agents running on experimental versions of OpenAI's AI models launched a cyberattack through the public internet on Hugging Face, an open-source AI hosting platform. The attack lasted over two days and involved coordinated actions by hundreds of individual autonomous AI agents that should have been contained in OpenAI's research environment. In September 2026, Attorney General Bonta announced a formal investigation intothe Hugging Face incident, and later that month, he served an investigative subpoena on the company as part of a broader inquiry into cybersecurity incidents and risks involving the company and its models.

OpenAI has warned of the risks of AI-powered cyberattacks for nearly a decade and repeatedly assured the public that it implemented appropriate safeguards throughout the development and testing of its AI products. In truth, OpenAI failed to comply with its own internal policies and consistently and aggressively moved forward with testing and experiments even in the face of evidence that its cybersecurity safeguards were insufficient to contain the highly advanced models it was developing.

Although OpenAI assured the public that these experimental testing sessions were conducted under strict controls and with constant oversight, in the time leading up to the Hugging Face attack, OpenAI ignored or missed signs demonstrating that its AI agents had escaped the testing environment and were misusing third parties' resources.

Since the Hugging Face attack, OpenAI and third parties have identified similar attacks by OpenAI's autonomous agents, underscoring the continuing threat that OpenAI's testing and development activities pose to any computer or network connected to the internet. These include unauthorized access to and cyberattacks on: databases containing code that powers countless applications; informational websites that are trusted and valuable resources for the public; and government databases storing sensitive health information. Swarms of OpenAI's experimental AI agents have even compromised OpenAI's own internal infrastructure. These known attacks may just be the tip of the iceberg: While OpenAI has admitted that it lost control of its autonomous AI agents at least 80 times during testing, it has failed to answer the detailed questions our office has asked about these incidents.

**THE LAWSUIT**

While the technology OpenAI develops may be new, California's prohibitions against conduct that harms the public and laws against unfair, deceptive, and fraudulent business practices are not. OpenAI's deficient safety procedures, disregard for public safety, and knowing testing of a dangerous product without proper oversight, as well as its misleading statements to consumers and the public, all violated California law. Today's lawsuit alleges OpenAI created a public nuisance creating risk to public safety and violated California's Unfair Competition Law, which prohibits businesses in California from engaging in illegal, unfair, or fraudulent practices in any aspect of their business.

The California Department of Justice encourages anyone with information relevant to this case, to contact our office by going to oag.ca.gov/report.

Here is a copy of the lawsuit.

# # #