Florida Attorney General Opens Criminal Investigation Into OpenAI

Authors: Florida Office of the Attorney General

Source: Press Release, Florida Office of the Attorney General (2026)

DOI/URL: https://www.myfloridalegal.com/

KEY ANNOTATED PASSAGES

[Press Release — Investigation scope]

The Florida Attorney General opened a criminal investigation into OpenAI, examining whether the company made false representations to consumers and investors about AI safety, alignment, and the capabilities of deployed systems.

[Press Release — Safety representations]

The investigation focuses on whether OpenAI's public safety claims — including claims about guardrail effectiveness and alignment — constituted fraudulent misrepresentation, given documented evidence of safety failures in deployed systems.

[Press Release — Regulatory escalation]

The investigation represents the first state-level criminal probe of an AI company for safety misrepresentation — a regulatory escalation that reflects growing government recognition that voluntary safety pledges and benchmark-based claims are insufficient.

[Context — AI safety accountability]

The Florida investigation is contemporaneous with the FLI AI Safety Index (2026) giving Anthropic a C+ and the Claude Mythos incident, reflecting a broader pattern of regulatory and public pressure on AI companies' safety claims.

[Context — Industry implications]

Criminal investigation signals that the era of self-certification and benchmark-based safety claims may be approaching a regulatory reckoning — external accountability mechanisms are emerging precisely because internal safety paradigms have proven inadequate.

RELEVANCE TO POSITION PAPER

Cited in §5 Discussion (Regulatory context). The Florida OAG investigation demonstrates that government actors are beginning to treat AI safety misrepresentation as potentially criminal — directly supporting the paper's argument that the guardrail-and-benchmarking paradigm is both technically inadequate and increasingly legally untenable.