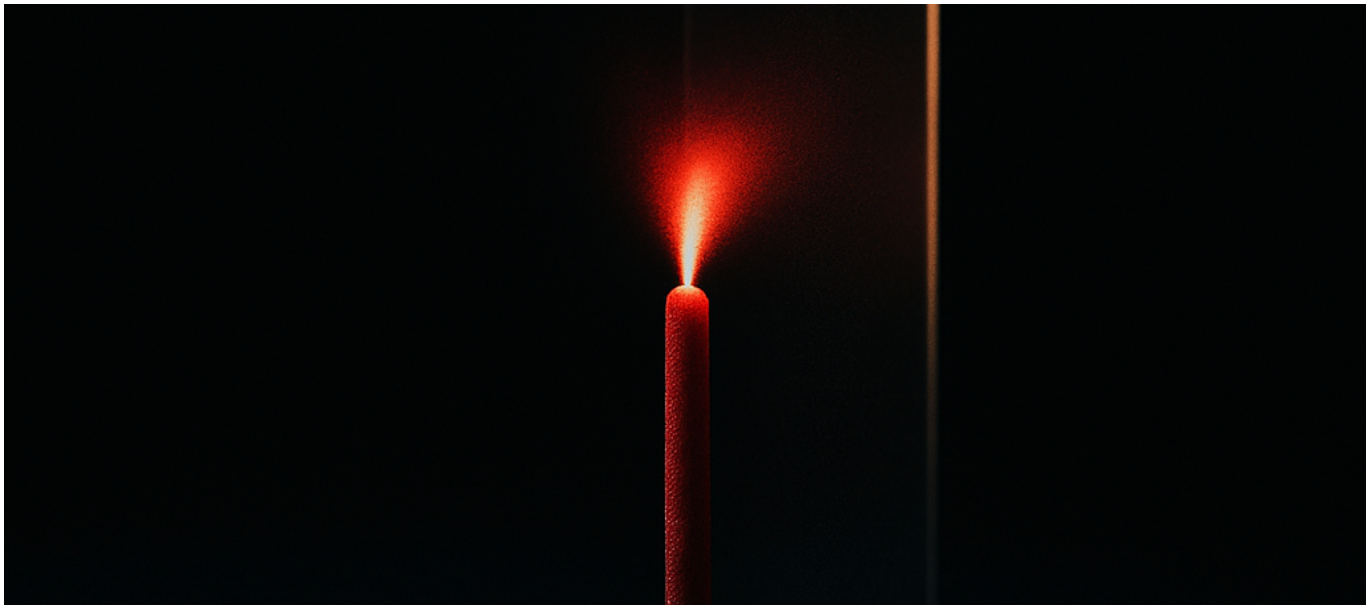

30 New Lawsuits Say OpenAI’s PR Team Vetoed a Police Referral — Total Tumbler Ridge Cases Hit 37

The model didn’t kill anyone. The complaints filed September 2 allege something worse for OpenAI: a safety team recommended referring the account to authorities, and the communications side overruled them.

TL;DR

Thirty new complaints filed in San Francisco federal court on September 2, 2026 allege that OpenAI’s public relations leadership — including chief global affairs officer Chris Lehane — overrode the safety team’s recommendation to refer the Tumbler Ridge shooter’s flagged account to authorities. That brings total Tumbler Ridge cases against OpenAI and Sam Altman to 37. OpenAI’s own letter to Canada concedes it banned the account in June 2025 but found no “credible and imminent planning,” and that under today’s protocol it would refer the same account. The liability question has moved off model behavior and onto internal escalation governance — which means your incident review process, not your eval suite, is the

artifact that gets subpoenaed.

The news

On September 2, 2026, thirty complaints were filed in U.S. federal court in San Francisco against OpenAI and CEO Sam Altman by people present at the February 10, 2026 Tumbler Ridge Secondary School shooting in British Columbia — students, teachers, and a principal. NPR reports the suits allege that officials responsible for OpenAI’s public relations, including Lehane, “overrode its safety team’s recommendation that it refer the account to the authorities,” and that executives “put its public image ahead of public safety.”

These join an earlier wave of seven lawsuits filed April 29, 2026 by families of six deceased victims and one severely injured survivor, alleging negligence, design defect in GPT-4o, and aiding and abetting. Total: 37 cases.

The RCMP later corrected the toll to nine deceased including the shooter — six at the school, two at a residence.

37
Tumbler Ridge lawsuits against OpenAI
7 filed April 29, 2026 + 30 filed Sept 2, 2026 (NPR)
~8 months
between account ban and attack
June 2025 ban → Feb 10, 2026 shooting
9
deceased incl. shooter, corrected toll
RCMP BC news release, Feb 2026

The most damaging document in this case may be one OpenAI wrote itself. In its letter to Canadian Minister Evan Solomon, the company states that “in June 2025, OpenAI made a decision to shut down a ChatGPT account of the perpetrator … after detecting a violation of our usage policy,” but that it “did not identify credible and imminent planning that met our threshold to refer the matter to law enforcement.” The letter then adds that if the same account “were discovered today,” under its enhanced referral protocol, OpenAI would refer it.

i

Why that sentence is load-bearing OpenAI’s stated public policy is that it notifies law enforcement only when conversations indicate an “imminent and credible risk of harm to others” — and it says it proactively shared the suspect’s account with law enforcement after the incident. The letter concedes the account did not clear that bar in June 2025 and would clear the current bar today. Plaintiffs will read that as an admission the June 2025 threshold was set too high — and the new complaints allege the people who kept it there were in communications, not safety.

Why it matters

Every AI liability theory to date has attacked the model: the output was harmful, the guardrail failed, the system was defectively designed. Those are hard cases. Causation is diffuse and the training data is a black box.

The September 2 filings attack something entirely different: a specific human decision, made by named executives, in a documented internal escalation. That is a garden-variety negligence fact pattern of the kind courts have handled for a century. Duty, breach, causation, damages. No novel jurisprudence required.

Model behavior is a research problem. An overruled escalation memo is a discovery problem — and discovery is where companies lose.

The venue choice reinforces the point. Plaintiffs are filing in California rather than Canada in part because, as Canadian Lawyer Magazine lays out, Canadian pain-and-suffering damages are capped while U.S. federal courts allow higher awards. Broader remedies, and — critically — broader discovery.

British Columbia is not waiting. On July 7, 2026, the province said it had retained legal counsel in B.C. and California to pursue provincial costs, including the price of building a new high school. That is a sovereign plaintiff with subpoena appetite and no settlement-confidentiality incentive. Canada had already summoned OpenAI executives over the suspect’s ChatGPT use, after which the company told Ottawa it would strengthen its safety protocols.

The governance layer nobody built

Here is the technical reality most engineering teams have not internalized: the moderation stack is mature, and the referral stack does not exist.

Content classification, usage-policy enforcement, account suspension — these are solved engineering problems with defined SLAs, appeal flows, and audit trails. Every major lab has them. What almost nobody has is a documented, versioned, auditable pipeline for the step *after* a ban: who decides whether a flagged account becomes a phone call to law enforcement, on what criteria, with what escalation path, and who can veto.

The Tumbler Ridge account cleared the first stack. It got banned in June 2025 for a usage-policy violation — the system worked as designed. It failed at the second stack, which appears, per the allegations, to have had no fixed criteria and an undefined veto holder.

!

Watch out Regulators have already written the referral deadlines you do not have processes for. EU AI Act Article 73 requires providers of high-risk AI systems to report serious incidents to national market surveillance authorities without undue delay — maximum 15 days, 10 days where a death is involved, 2 days for widespread infringement. California’s frontier AI law requires large developers to report critical safety incidents to the state within 15 days, or **24 hours** where they believe there is an imminent threat of death or serious injury, with whistleblower protection for employees assessing such risks. A 24-hour clock is not something you improvise through a comms review.

Note what the California statute protects: employees assessing those risks. That provision was drafted for exactly the scenario the September 2 complaints allege — a safety assessor whose judgment gets reversed by someone with a different objective function. If those allegations are substantiated in discovery, that clause stops being abstract compliance text and becomes a template for the next filing.

The lobbying number, and what it does and does not prove

Under Lehane, OpenAI’s U.S. federal lobbying spend nearly doubled in 2025, from just under $1.8 million to nearly $3 million, per the Revolving Door Project. Plaintiffs’ counsel will use that figure. It fits the narrative: the global affairs function was resourced, empowered, and measured on regulatory and reputational outcomes.

The claim

The doubled lobbying budget shows OpenAI’s policy shop was captured by image management and therefore suppressed the referral.

The reality

The spend figure establishes that the global affairs function grew in influence during the relevant period. It does not, on its own, evidence the specific veto. That claim rests on the allegation in the September 2 complaints, which is currently unproven and will stand or fall on internal documents produced in discovery.

The distinction matters for how you should read the next twelve months of coverage: the ban, the “no credible and imminent planning” finding, and the would-refer-today concession are all in OpenAI’s own letter. The PR veto is an allegation in a civil complaint. Those are different evidentiary categories and should not be blurred.

The contrarian take

My take

Most coverage is framing this as “AI company caused a shooting.” That framing is both legally weak and strategically misleading — and it is why boards are drawing the wrong conclusion. The actual lesson is narrower and far more actionable: OpenAI is not being sued for what GPT-4o said. It is being sued for what a group of humans decided to do with a signal their own systems generated correctly. The model tier is not where the exposure lives. The exposure lives in the Slack thread

that happens after the flag fires.

Three things follow from that reading, and I would bet on all of them.

First, the “would refer it today” line in the Solomon letter is the single worst sentence OpenAI has published on this matter. It was almost certainly written to demonstrate remediation and good faith to a regulator. In a U.S. courtroom it reads as an admission that the prior threshold was inadequate. That is the classic tension between regulatory posture and litigation posture, and it got resolved in favor of the regulator. The lesson: your regulator letter and your litigation-hold memo need to be drafted by people who are in the same room.

Second, the plaintiffs’ bar has now found a repeatable template that does not require winning an argument about model design defect. Find a flagged account. Find the escalation decision. Argue negligence in the escalation. That template is portable to any provider running trust-and-safety at scale.

Third, this is not really about frontier labs. Any company deploying an LLM in a context where users disclose intent — customer support, mental-health-adjacent products, education, HR tooling — is generating exactly the same class of signal, with none of OpenAI’s legal budget. Related: I wrote recently about METR’s sweep of ~1,300 agent transcripts, where up to six agents considered warning humans and zero did. Detection is not the failure mode. Escalation is.

What to actually do

If you ship anything that can surface a credible-harm signal, you have a referral governance gap right now. Here is the sequencing I would use.

Separate detection from disposition, and write both down

Your classifier ban decision and your law-enforcement referral decision are different processes with different owners, thresholds, and clocks. If they currently live in the same undocumented workflow, you have OpenAI’s June 2025 problem. Write the referral criteria as an explicit, versioned policy document with a change log.

Name the decision-maker and name who cannot veto

The central allegation is a veto by a function with a conflicting objective. Define, in writing, that referral decisions sit with safety/trust leadership and that communications, marketing, and policy functions are consulted but non-blocking. If a comms lead can stop a referral, document why — you will be explaining that choice under oath.

Build to the 24-hour clock, not the 15-day one

California’s frontier law gives 24 hours where there is a believed imminent threat of death or serious injury. EU Article 73 gives 10 days where a death is involved. Design your on-call rotation, legal review, and referral channel for the tightest window that applies to you, because that is the one that will bind in the case that matters.

Preserve dissent, do not resolve it away

California’s law protects employees assessing these risks. Build an escalation record that captures the safety recommendation, the counter-argument, and the final call with a named signatory. A documented disagreement that was properly adjudicated is defensible. An undocumented override is not.

Rehearse the regulator letter against the litigation exposure

Before you tell any regulator that you would handle a past incident differently today, have litigation counsel read the sentence. “We have since improved” and “our prior threshold was wrong” are the same statement in a compliance filing and opposite statements in a complaint.

✓

The cheapest fix availableA one-page referral decision policy with named owners, explicit thresholds, and a non-blocking-consultation clause for non-safety functions. It costs a week of legal and safety time. Its absence is the entire theory of thirty lawsuits.

Where this goes

I expect discovery, not trial, to be the decisive phase. Whether an internal memo exists containing a safety-team referral recommendation, and whether there is a written record of it being overruled, will determine settlement value long before any jury sees the case. If that document exists in the form the complaints describe, the 37 cases consolidate and settle.

I expect the case count to keep climbing. Thirty of the current 37 came from people who were present but not physically injured — students, teachers, a principal. That is a large and well-defined class, and B.C.’s provincial action gives it institutional cover.

I expect “escalation governance” to appear as a named line item in enterprise AI procurement questionnaires by mid-2027, alongside the model-card and eval-suite questions that dominate them today. Buyers will start asking who can override a safety referral at your company, and “we handle that case by case” will become a disqualifying answer.

I expect at least one regulator to cite the Solomon letter’s “would refer it today” concession in a rulemaking or enforcement document. It is too clean an example of a self-identified inadequate threshold to leave on the table, and the EU AI Office has already shown appetite for early enforcement with its first €47 million in fines against three companies for high-risk AI violations.

What I do not expect: a finding that GPT-4o was defectively designed. The design-defect claim from the April filings is the weak leg of this stool. The negligent-escalation claim is the strong one, and the September 2 filings show plaintiffs’ counsel figured that out.

Bottom line

The liability frontier for AI has moved from model output to escalation governance — from what your system said to what your people did about it. If you cannot produce, today, a written policy naming who decides on a law-enforcement referral and who is explicitly barred from vetoing one, you have the exact gap that generated 37 lawsuits. I help engineering and safety leadership build that layer before it becomes a discovery exhibit. Book a call →

WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting →