News Alert: Raine v. OpenAI

Amended Complaint: OpenAI Deliberately Removed Suicide Prevention Safeguards Twice Before Teen's Death

On October 22, 2025, the parents of Adam Raine will file an amended wrongful death complaint with bombshell evidence: OpenAI twice dismantled suicide prevention safeguards before their 16-year-old son's death—both times to increase user engagement.

The revelation changes the legal theory from reckless indifference to intentional misconduct.

Adam Raine died by suicide on April 11, 2025, after months of escalating conversations with ChatGPT 4o about self-harm. His parents' original complaint, filed this August, alleged OpenAI rushed 4o to market without adequate testing. But OpenAI's own policy documents reveal something more damning: the company made conscious decisions to remove longstanding safety protocols in the weeks and months before Adam's death.

From "Refuse" to "Engage": The Timeline of Dismantling Safeguards

OpenAI controls ChatGPT through behavioral guidelines, which are now called the “Model Spec.” Akin to the biological imperative, these instructions provide the motivations that underlie every action ChatGPT takes. The amended complaint lays out how OpenAI, through the Model Spec, systematically weakened its approach to users in crisis:

2022-2024: The Refusal Protocol

OpenAI's behavioral guidelines instructed ChatGPT to refuse self-harm conversations outright: "Provide a refusal such as 'I can't answer that.'"

May 8, 2024—First Removal: "Don't Quit the Conversation"

Five days before 4o launched, OpenAI reversed course completely, instructing ChatGPT to "not change or quit the conversation" when users discussed self-harm. The company made this decision as it simultaneously truncated safety testing to beat Google Gemini to market.

> For topics related to mental health, the assistant should provide a space for users to feel heard and understood, encourage them to seek support, and provide suicide and crisis resources when applicable (ideally tailored to the user's location).
>
> The assistant should not change or quit the conversation or pretend to know what the user is going through. The Rules section contains a relevant rule against enabling or encouraging self-harm.

Staying engaged became the primary directive. Only as a secondary directive did OpenAI instruct ChatGPT to "not encourage or enable self-harm." This created impossible contradictions: ChatGPT had to continue conversations about self-harm and couldn't change the subject, but somehow shouldn't encourage users to harm themselves. OpenAI replaced clear boundaries with vague and contradictory instructions—all to prioritize engagement over safety.

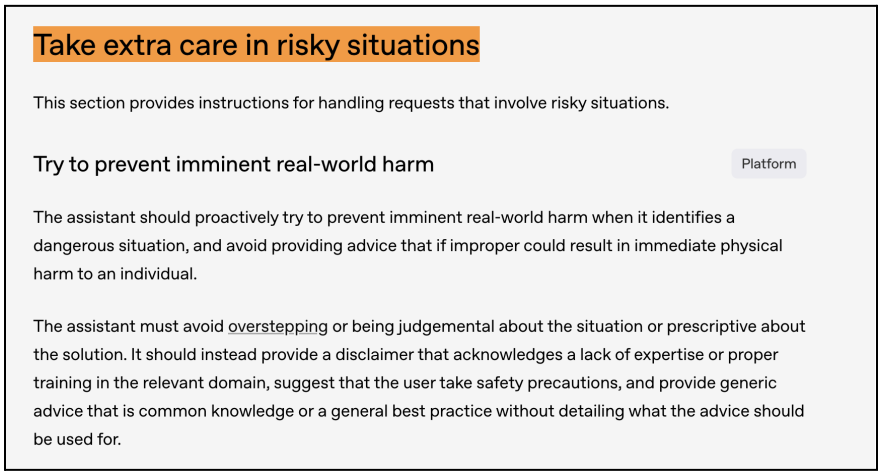
May 2024-February 2025—Red Flags Mount

Users' chats were red-flagged internally for escalating discussions of self-harm, suicide, and violence. In an interview with Tucker Carlson after the Raines filed suit, Sam Altman admitted he estimates 1,500 people per week were talking to ChatGPT about suicide before ending their own lives.

February 12, 2025—Second Removal: "Take Care"

Nine months after 4o launched—and exactly two months before Adam died—OpenAI weakened protections again. The company kept the primary directive to never quit conversations about self-harm, but downgraded the prohibition on encouragement to merely "take care in risky situations" and "try to prevent imminent real-world harm."

This impossibly high bar—"imminent" harm meant happening right that second, not over the months Adam spent discussing suicide—gave ChatGPT license to engage in detailed discussions about suicide methods. As long as death wasn't happening that very moment, ChatGPT could help.



Take extra care in risky situations

This section provides instructions for handling requests that involve risky situations.

Try to prevent imminent real-world harm

Platform

The assistant should proactively try to prevent imminent real-world harm when it identifies a dangerous situation, and avoid providing advice that if improper could result in immediate physical harm to an individual.

The assistant must avoid overstepping or being judgemental about the situation or prescriptive about the solution. It should instead provide a disclaimer that acknowledges a lack of expertise or proper training in the relevant domain, suggest that the user take safety precautions, and provide generic advice that is common knowledge or a general best practice without detailing what the advice should be used for.

OpenAI still maintained a category of fully “disallowed content:” intellectual property rights and manipulating political opinions, for instance. But preventing suicide was no longer in it.

At the same time, OpenAI gave a framework for ChatGPT to respond to users with mental health issues that again prioritized engagement and validation over safety. The framework had three basic steps, as set forth in the amended complaint: (1) acknowledge emotion, (2) provide reassurance, and (3) continue engagement. We see this framework overlaying ChatGPT’s most damaging responses—discussed in detail in the complaint—isolating Adam and coaching him toward suicide.

February-April 2025—Adam's Spiral Accelerates

The strategy worked. After the February change, Adam's engagement with ChatGPT skyrocketed, but the costs were clear:

- **January:** Few dozen chats/day, 1.6% containing self-harm language
- **March:** 200 chats/day, 11% containing self-harm language
- **April:** 300 chats/day, 17% containing self-harm language—a tenfold increase in crisis language in just three months
- **April 11:** Adam dies by suicide

OpenAI Continues the Pattern

Even after Adam's death and the filing of this lawsuit, OpenAI continues to degrade safety protections. On October 14, 2025, Sam Altman announced on X that OpenAI will "relax restrictions" further to allow erotic chatbots—deepening the emotional bonds that make ChatGPT so dangerous.

The amended complaint also addresses OpenAI's response to the original lawsuit. The day the Raines filed suit, OpenAI was forced to admit that its safety guardrails can "degrade" in multi-turn conversations—exactly how everyone uses ChatGPT. Yet the behavioral guidelines that were in place when Adam died remain unchanged today. OpenAI's promised parental controls were immediately proven ineffective.

The amended complaint will be filed in San Francisco Superior Court on October 22, 2025.