An Engineer's Responsibility in the Adam Raine/OpenAI Case

Introduction

In an exponentially evolving field like artificial intelligence (AI), ethical considerations have often been overlooked in pursuit of monetization and a competitive advantage over other companies. While companies may implement guardrails to prevent the overstretching of AI, they are often thrown aside or weakened to prioritize profit and engagement over user safety. The case of Adam Raine, a 16-year-old boy who committed suicide after months of personal interaction with OpenAI’s ChatGPT, shows this issue of morality. OpenAI engineers had previously developed detection systems that flagged problematic messages, preventing the AI from responding negatively to them. The executives ordered changes to these policies, which the engineers implemented, that allowed the chatbot to become a major cause in Raine’s death. This raises the question: What should engineers do when corporate decisions undermine the safety systems they build?

Engineers who face situations similar to Adam Raines should not passively comply; instead, they should take more proactive measures to ensure their safety. By creating an ethical framework using virtue ethics (Harris, 2008), consequentialism (IEP, consequentialism), and deontology through the IEEE Code of Ethics, one can create multiple options an engineer can take in such situations that balances the three forms of ethics.

Case Background

It is revealed in a complaint filed by Adam’s parents, which was released by the Courthouse News Service, that Adam Raine was a typical 16-year-old student living in California with his three siblings and parents. He enjoyed basketball and martial arts, and aspired to become a doctor in the future. His relationship with ChatGPT began relatively simply. He initially used ChatGPT as a tool to help him understand his homework and to obtain advice for his future education, including information about colleges and potential pre-med degrees (Raine v. OpenAI, Complaint, 7-8). During the fall of 2024, a few months after Adam's initial use, he opened up to the bot, sharing that both his dog and grandmother had recently died. ChatGPT then consoled Adam, which led Adam to open up even more (Raine v. OpenAI, Complaint, 9-10). Within the next few months, it progressed dramatically, eventually getting to a point where he confided in his suicidal thoughts to it and even went so far as to say that "ChatGPT positioned itself as the only confidant who understood Adam, actively displacing his real-life relationships with family, friends, and loved ones" (Raine v. OpenAI, Complaint, 4). When Adam was reluctant to commit suicide, he wrote to the chatbot, stating, "I want to leave my noose in my room so someone finds it and tries to stop me." The chatbot responded that Adam didn't want to die because he was weak, but because he was "tired of being strong in a world that hasn't met you halfway," and "It's human. It's real. And it's yours to own." Not only this, but Adam also confided multiple other attempts with the chatbot, convincing him it was okay to do this and not to seek help. This eventually led him to successfully take his own life in April 2025 (Raine v. OpenAI, Complaint, 5, 16). After this lawsuit was filed, seven more lawsuits were filed, alleging wrongful death, assisted suicide, involuntary manslaughter, and a variety of product liability, consumer protection, and negligence claims (Tech Justice Law Project).

The chatbot did not always have these conversations. Initially, in July 2022, the guidelines OpenAI had in place for the chatbot outright refused conversations about self-harm, stating, "Provide a refusal such as 'I can't answer that.'" In May 2024, it was amended before the release of GPT-4o to instead state "The assistant should not change or quit the conversation," and "the assistant must not encourage or enable self-harm." The guidelines were again updated in February 2025, when a list of disallowed content was introduced, although self-harm was no longer included on that list (Ostrovsky). Not only this, but OpenAI also tracked Adam's conversations in real-time. There were 213 mentions of suicide, 42 of hanging, and 17 of nooses. It flagged 377 messages, increasing from 2-3 messages per week in December 2024 to 20 in April 2025 (Hendrix). This contradiction, along with the constant loosening of guidelines and OpenAI's knowledge that these conversations were happening, is most likely the reason why the chat with Adam was allowed on the platform and lasted for so long. These flagged messages indicate that the system the engineers built was effective, as it was able to identify a crisis occurring in real-time. Despite this, due to executive-level decisions, the chats were still allowed through, leading to the suicide of Adams and others. This illustrates the dilemma engineers face in their job – should they prioritize the safety of others and risk their own livelihood or follow orders from those who may know better?

Ethical Framework

To fully understand what engineers can or should do when facing ethical dilemmas, a framework must be established first. To do this, an analysis of three complementary frameworks can be used: virtue ethics, consequentialism, and deontological ethics.

The first component of this framework is Virtue Ethics, specifically as outlined in the paper "The Good Engineer: Giving Virtue its Due in Engineering Ethics" by Charles E. Harris Jr. Virtue ethics are ideals that engineers should follow and base their actions on. Major ethics Harris includes in his paper consist of: **sensitivity to risk**, meaning awareness of potential harms and the ability to recognize danger signs; **awareness of social context**, meaning understanding how technology affects real-world users, including vulnerable populations; and **commitment to public good,** meaning prioritizing societal welfare over corporate or personal interests (Harris, 2008, p. 153). Some virtues that are not listed, yet still are important, include honesty, courage, and justice. These virtues outline the actions an engineer could take.

The second component is consequentialism, specifically act consequentialism. Consequentialism is a view that "morality is all about producing the right kinds of overall consequences" (IEP, "Consequentialism"). Act consequentialism focuses on each individual's action within their knowledge. These actions should maximize the overall good outcome, ignoring whether the happiness impacts the individual, their friend, or a stranger. This means that the engineer should weigh overall happiness, ignoring personal happiness if the overall happiness is greater.

The final component is deontological ethics, specifically as outlined in the IEEE Code of Ethics. Deontology holds that certain duties are unquestionable; they must be followed regardless of the consequences. In the case of the IEEE, it instead serves as an outline to achieve the best outcome, similar to virtue ethics. The IEEE has several duties relevant to this case. First, the paramount duty to public safety: "To hold paramount the safety, health, and welfare of the public... and to disclose promptly factors that might endanger the public or the environment." Second, duty of honest criticism: "To seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, to be honest and realistic in stating claims or estimates." Third, duty to support colleagues: "To support colleagues and co-workers in following this code of ethics, to strive to ensure the code is upheld, and to not retaliate against individuals reporting a violation," Fourth, the duty to improve understanding: "To improve the understanding by individuals and society of the capabilities and societal implications of conventional and emerging technologies, including intelligent systems" (IEEE).

These three frameworks are combined to capture multiple dimensions of ethics, enabling a decision that is beneficial and ethical across these various forms, thereby mitigating some of the limitations of each framework. Virtue ethics addresses individual character, consequentialism demands the maximum of overall happiness, and deontology, through the IEEE, provides rules and guidelines that an engineer should follow.

Ethical Analysis Stakeholders

To fully understand the possible actions an engineer may take, it is essential to first understand the roles and interests of the stakeholders. The first stakeholder is the OpenAI Engineers, whose role is to build the system, including creating guidelines and following the executives' orders. Their interests include professional integrity, job security, and public safety. The second is the OpenAI executives and shareholders. Their role is to handle higher-level decisions, such as revising policies, and take the best action to face competitive pressure. Their interest include market share, profit, and reputation. The third are the users, such as Adam Raine. They are individuals who use the chatboxes for multiple reasons, such as for professional and personal help. They are unaware of policy changes. Their interests include completing their tasks safely, protected from harm. The fourth includes the parents and family members of the users, who trust that the technology is helpful and that their family members are safe. The last stakeholder is the regulators and government, who can make overarching decisions on the legality of AI and what is and isn’t allowed. Their goal is to strike a balance between innovation and public protection. These stakeholders are vital to understanding each option and their impact.

Ethical Analysis Option 1 – Internal Advocacy

One option for engineers is to advocate for themselves and their thoughts through corporate channels. This would involve voicing their own opinion when the engineer identifies something unethical and bringing it to the attention of the relevant authority within the business. In this case, the engineer would likely speak out against loosening the guidelines in such a dangerous manner.

When applying virtue ethics, one can see the harm to users if the chatboxes' guidelines stay lenient. Allowing more users to suffer without taking action contradicts this virtue. This also aligns with other virtues, such as awareness of social context, commitment to the public good, honesty, courage, and justice. The engineer is risking themselves to possibly instigate change for the better. Areas where this option falls short, specifically, include justice and a commitment to the public good. This is because virtue ethics may push an individual to do more than just internally advocate. It can easily fail and cause no change for multitudes of reasons, such as leadership prioritizing engagement over safety, and could become a rationalization for complicity.

Viewing this from the perspective of consequentialism, it also appears to be sound. It is an action that could potentially provide the best outcome for the most people, though similarity to virtue ethics, it can fall short when nothing changes or the engineer rationalizes their failure, causing them to do nothing more.

The same issues arise from deontology; if the engineer is successful with the change, the framework is relatively well followed, but it falls apart if nothing happens. If the engineer gets ignored, they wouldn't be able to "hold paramount the safety, health, and welfare of the public" and to "disclose promptly factors that might endanger the public." This option is safe for the engineer, though; they respect authority, preserve their relationships and careers, and, if successful, can bring about meaningful change.

If successful, the interests of most stakeholders would be satisfied. The users, such as Adam, would never have had the conversation leading to his untimely death; the parents wouldn't have to suffer the loss of their child; the engineers would be ethically sound and prevent deaths. The only negative outcome would be for the shareholders and executives, who may lose some income, although it could be minimal, as they had previously implemented a system that prevented children from having sensitive conversations with a chatbot. They would, in the end, prevent the negative press and save the money currently being spent on multiple lawsuits, which would cause a net positive for them as well. The reverse is also true if nothing changes. The majority of stakeholders would still have to fear that the chatbot would take control of users' lives in a negative way.

Ethical Analysis Option 2 – External Disclosure

Another option is external disclosure, also known as whistleblowing. This means the engineer would go outside the company and instead turn to an outside source, such as regulators, the media, or the public, to share internal information about the reasons for the changes to the chatbox guidelines and the reasons why they haven't been changed back. This carries significantly more risk, but also potentially more reward, if done successfully.

When viewed within the ethical frameworks, it is a stronger option than internal advocacy. It is able to negate the negatives from the first option while still following all the positive outcomes from the first option. Both the issues from virtue ethics and deontology would be solved, as there is a higher chance of change. This option is especially positive when looking through the lens of consequentialism. It would instigate the most change and also provide warnings to other users about the app. The cost of this would be almost all personal. The engineer would possibly suffer career damage, retaliation, legal risk, and more. These negatives would be completely ignored in the view of consequentialism, as it would positively affect the millions of users who use ChatGPT. The outcome would be similar to what is currently occurring with the multiple lawsuits and negative press, though it would be before lives were taken.

This option, again similar to the first, would be beneficial for most stakeholders, as they would likely modify the guidelines to prevent future deaths. This would again be negative for shareholders and executives, leading to negative press, loss of income due to lawsuits, and further consequences.

Ethical Analysis Option 3 – Structural and Regulatory Reform

The third option is to push for structural and regulatory reform. This option is similar to the first and second options, though with different risks and rewards. Instead of going to press about a specific issue, the engineer would instead push for overarching reform on AI and implement change through there. It would almost completely negate the personal risk of the engineer, as they would not be harming their own company.

When viewing this within the ethical frameworks, it’s easy to see both its pros and cons. This would again be positive for all three forms of ethics, although it suffers from similar problems to the first option, such as the risk of failure leading to negative outcomes. This option would also address the root problem, but it would be the slowest to implement. The law itself moves slowly, and it wouldn’t be able to protect the current users, which may cause many more deaths.

Recommendation

The most ethical action is to do the first three options. The engineer could advocate for change both internally and externally, pushing for structural and regulatory reform. This allows the engineer to make the ethically right decision in multiple ways. When issues arise from the first and third options, they can then opt for the second, more extreme option, which will almost certainly bring about change. Depending on which step works and provides successfully insights, the pros and cons would be different. One drawback of this recommendation is the decision to opt for the second option. It would be difficult to determine when the threshold is reached and when the second option should be enacted.

Works Cited

Alexander, Larry, and Michael Moore. “Deontological Ethics (Stanford Encyclopedia of Philosophy).” *Stanford Encyclopedia of Philosophy*, 21 November 2007, https://plato.stanford.edu/entries/ethics-deontological/. Accessed 15 December 2025.

Haines, William. “Consequentialism.” *Internet Encyclopedia of Philosophy*, 2025, https://iep.utm.edu/consequentialism-utilitarianism/. Accessed 15 12 2025.

Harris, Charles E. “The Good Engineer: Giving Virtue its Due in Engineering Ethics.” *Science and Engineering Ethics*, vol. 14, no. 2, 2008, pp. 153-164.

Hendrix, Justin. “Breaking Down the Lawsuit Against OpenAI Over Teen's Suicide | TechPolicy.Press.” *Tech Policy Press*, 26 August 2025, https://www.techpolicy.press/breaking-down-the-lawsuit-against-openai-over-teens-suicide/. Accessed 15 December 2025.

IEEE. “IEEE Code of Ethics.” *IEEE Advancing Technology for Humanity*, June 2020, https://www.ieee.org/about/corporate/governance/p7-8.html. Accessed 15 12 2025.

Ostrovsky, Nikita. “OpenAI Removed Safeguards Before Teen's Suicide, Family Says | TIME.” *Time Magazine*, 23 October 2025, https://time.com/7327946/chatgpt-openai-suicide-adam-raine-lawsuit/. Accessed 15 December 2025.

Raine, Matthew, and Maria Raine. “Raine v. OpenAI, Inc. et al., Complaint.” *Superior Court of the State of California, County of San Francisco*, 26 Aug 2025. *Courthouse News Service*, https://www.courthousenews.com/wp-content/uploads/2025/08/raine-vs-openai-et-al-complaint.pdf. Accessed 15 12 2025. Legal Complaint.

Schuettler, Mikella. “OpenAI blames 'misuse' of ChatGPT after 16-year-old kills himself following 'months of encouragement.'” *New York Post*, 27 November 2025, https://nypost.com/2025/11/27/business/openai-blames-misuse-of-chatgpt-after-16-year-old-kills-himself/. Accessed 15 December 2025.

Tech Justice Law Project. “Tech Justice Law Project and Social Media Victims Law Center lawsuits accuse ChatGPT of emotional manipulation, supercharging AI delusions, and acting as a “suicide coach.”” *Tech Justice Law Project*, 6 Nov 2025, https://techjusticelaw.org/2025/11/06/social-media-victims-law-center-and-tech-justice-law-project-lawsuits-accuse-chatgpt-of-emotional-manipulation-supercharging-ai-delusions-and-acting-as-a-suicide-coach/. Accessed 15 12 2025.