PUBLICCITIZEN
1600 20th Street, NW • Washington, D.C. 20009 • 202/588-1000 • www.citizen.org

June 6, 2024

Attorney General Rob Bonta
Office of the Attorney General
California Department of Justice
1300 I Street
Sacramento, CA 95814

Christopher Lamerdin
Deputy Attorney General
Office of the Attorney General
Charitable Trusts Section
1300 I Street
Sacramento, CA 95814

RE: OpenAI, Inc.

Dear Attorney General Bonta and Deputy Attorney General Lamerdin:

This letter is to follow up on my letters of January 9 and March 5 encouraging you to investigate OpenAI, Inc.'s charitable status. In those letters, I reviewed a number of developments which supported the conclusion that OpenAI is no longer serving its public, nonprofit purpose and is instead effectively controlled by the for-profit OpenAI affiliate. If OpenAI, Inc. is no longer serving its public, nonprofit purpose, then under California law it should be dissolved, with the value of its assets transferred to another charitable enterprise, such as one or more foundations devoted to artificial intelligence ethics and safety.

I am writing today with additional supplementary information supporting the theory that OpenAI has abandoned its nonprofit mission.

1. OpenAI's Safety Leaders Depart and OpenAI Closes its Safety Team

On May 14, OpenAI's safety leaders resigned: Ilya Sutskever[1] was one of OpenAI's founders and co-lead of OpenAI's long-term safety ("superalignment") team; Jan Leike[2] was the other

[1] Ilya Sutskever (@ilyasut), "After almost a decade, I have made the decision to leave OpenAI ..." X, May 14, 2024. https://x.com/ilyasut/status/1790517455628198322

[2] Jan Leike (@janleike), "Yesterday was my last day as head of alignment, superalignment lead, and executive @OpenAI." X, May 17, 2024. https://x.com/janleike/status/1791498174659715494

long-term safety team co-lead. When that team was formed in July of 2023, OpenAI said it would be provided with 20 percent of the compute power that the company had available.[3]

In a series of posts on X (formerly Twitter), Leike expressed sorrow about leaving OpenAI. But he said he felt compelled to leave for reasons that speak directly to the issue of OpenAI's nonprofit status. In short, he explained, OpenAI has subordinated safety to profit-seeking:

> I joined because I thought OpenAI would be the best place in the world to do this [safety] research. However, I have been disagreeing with OpenAI leadership about the company's core priorities for quite some time, until we finally reached a breaking point.
>
> I believe much more of our bandwidth should be spent getting ready for the next generations of models, on security, monitoring, preparedness, safety, adversarial robustness, (super)alignment, confidentiality, societal impact, and related topics.
>
> These problems are quite hard to get right, and I am concerned we aren't on a trajectory to get there.
>
> Over the past few months my team has been sailing against the wind. Sometimes we were struggling for compute and it was getting harder and harder to get this crucial research done.
>
> Building smarter-than-human machines is an inherently dangerous endeavor. OpenAI is shouldering an enormous responsibility on behalf of all of humanity.
>
> But over the past years, safety culture and processes have taken a backseat to shiny products.

Leike doesn't just say he had disagreements with OpenAI's leadership. He specifies that OpenAI was depriving the long-term safety team of the resources needed to do its job. And he offers the very damning conclusion that "safety culture and processes have taken a backseat to shiny products."

Coming from a departed safety team leader, this would be – and is – a troubling assessment about any cutting-edge AI company. But OpenAI is not any company. While it operates as a for-profit, it is supposed to operate under the direction of a nonprofit board charged with prioritizing safety and especially long-term safety.

As Sutskever and Leike left OpenAI, the company disbanded its long-term safety team altogether.[4] Coming less than a year after the high-profile team had been formed, this development was astounding.

---

[3] Jan Leike and Ilya Sutskever, "Introducing Superalignment," OpenAI, July 5, 2023. https://openai.com/index/introducing-superalignment/

[4] Will Knight, "OpenAI's Long-Term AI Risk Team Has Disbanded," WIRED, May 17, 2024. https://www.wired.com/story/openai-superalignment-team-disbanded/

In the wake of the Sutskever and Leike's departure and the disbanding of the superalignment team, OpenAI on May 28 announced the formation of a board level safety committee, charged with making new safety recommendations for OpenAI's frontier models.[5] It is hard to see this as anything more than a PR effort. The new committee consists of members of OpenAI's board, who do not bring the same AI safety and ethics expertise as the departed experts, plus CEO Sam Altman, and some internal staff. By contrast, prior to the November 2023 board shake-up, the board included experts on AI safety and ethics.

On June 4, a group of current and former OpenAI employees ratified these concerns in an open letter raising safety concerns and urging AI companies to protect whistleblowers and abandon restrictive non-disclosure agreements.[6]

As regards the core question of whether OpenAI is still functionally operating as a nonprofit, recall that long-term safety is a central pillar of the nonprofit's charter.[7] It is very hard to square the recent safety developments at OpenAI with the idea that the OpenAI nonprofit board is in fact prioritizing its nonprofit charter over for-profit considerations.

2. OpenAI Prepares to Release its Human-Sounding Voice Mode Feature

In May, OpenAI introduced GPT-4o, with plans to give widespread public access to its Voice Mode feature.[8] Voice Mode is an audio version of ChatGPT that communicates in convincingly human-sounding voices, demonstrates what appears to be emotion, displays what appears to be empathy and human expressiveness and communicates in casual, colloquial and often humorous terms. Voice Mode may be the most extreme example in existence of AI anthropomorphism.

There is no doubt that Voice Mode is extremely impressive from a consumer standpoint. There is also no doubt that adopting deceptively anthropomorphic design poses enormous social risks, as extensive research details.[9] The dangers include include: enabling fraudulent and unfair business practices; facilitating deep privacy intrusions as people volunteer information to an AI that

---

[5] "OpenAI Board Forms Safety and Security Committee," OpenAI, May 28, 2024. https://openai.com/index/openai-board-forms-safety-and-security-committee/

[6] Jacob Hilton and other current and former OpenAI employees, "A Right to Warn about Advanced Artificial Intelligence," June 4, 2024, https://righttowarn.ai. In a series of posts on X, one of those former employees, Carroll Wainwright, stated that his faith in the non-profit structure of OpenAI has "significantly waned." He wrote, "I worry that the board will not be able to effectively control the for-profit subsidiary, and I worry that the for-profit subsidiary will not be able to effectively prioritize the mission when the incentive to maximize profits is so strong." (Carroll Wainwright (@clwainwright), "Last week was my final week working at OpenAI…" X, June 4, 2024. https://x.com/clwainwright/status/1798013325495963655)

[7] "OpenAI Charter," OpenAI, 2018. https://openai.com/charter/

[8] "How the voices for ChatGPT were chosen," OpenAI, May 19, 2024. https://openai.com/index/how-the-voices-for-chatgpt-were-chosen/

[9] Iason Gabriel et al, "The Ethics of Advanced AI Assistants," Google DeepMind, April 19, 2024. https://arxiv.org/pdf/2404.16244 and Rick Claypool, "Chatbots Are Not People: Designed-In Dangers of Human-Like A.I. Systems," Public Citizen, September 26, 2023. https://www.citizen.org/article/chatbots-are-not-people-dangerous-human-like-anthropomorphic-ai-report/

sounds like a human; and promoting dangerous emotional dependence on AI assistants/companions.[10]

There are other, even more profound risks presented by Voice Mode, many highlighted in an important Google DeepMind paper.[11] These include fundamentally degrading human-human relationships; undermining humans' ability to accept different points of view and severely worsening social atomization; and deepening social dissatisfaction.

It is noteworthy that Google – a for-profit company that is not governed by a non-profit with a safety and alignment mandate – affirmatively decided not to adopt anthropomorphic voices for its AI assistant.[12]

By itself, OpenAI's decision to embrace AI anthropomorphism and undertake a reckless social experiment does not prove it is no longer operating with nonprofit purpose. But the decision does contradict the safety-first mission of the nonprofit and adds more weight to the already strong claim that OpenAI has abandoned its nonprofit mission.

3. OpenAI's Nonprofit Board Did Not Know About the Release of ChatGPT 3.0 and Other Information About the November OpenAI Shake Up

In announcing its decision to fire OpenAI CEO Sam Altman in November 2023, the OpenAI nonprofit board had cited Altman's alleged lack of candor.

One of the board members who was part of the decision was Helen Toner, a researcher at Georgetown University. Toner left the board when Altman was reinstated. Toner recently appeared on the TED AI show podcast,[13] in which she provided context for the board's decision to fire Altman.

In the TED AI podcast, Toner provides several specific claims relating to Altman's lack of candor. First, and most startlingly, she said, "When ChatGPT came out in November 2022, the board was not informed in advance of that. We learned about ChatGPT on Twitter."

Second, she said, "Sam didn't inform the board that he owned the OpenAI Startup Fund, even though he constantly was claiming to be an independent board member with no financial interest in the company."

---

[10] Rick Claypool, "Chatbots Are Not People: Designed-In Dangers of Human-Like AI Systems," Public Citizen, September 26, 2023, https://www.citizen.org/article/chatbots-are-not-people-dangerous-human-like-anthropomorphic-ai-report

[11] Iason Gabriel et al, "The Ethics of Advanced AI Assistants," Google DeepMind, April 19, 2024. https://arxiv.org/pdf/2404.16244

[12] Will Knight, "Prepare to Get Manipulated by Emotionally Expressive Chatbots," WIRED, May 15, 2024. https://www.wired.com/story/prepare-to-get-manipulated-by-emotionally-expressive-chatbots/

[13] "What really went down at OpenAI and the future of regulation w/ Helen Toner," The TED AI Show, May 2024. https://www.ted.com/talks/the_ted_ai_show_what_really_went_down_at_openai_and_the_future_of_regulation_w_helen_toner

Third, and more generally, she said, "On multiple occasions, he gave us inaccurate information about the small number of formal safety processes that the company did have in place – meaning that it was basically impossible for the board to know how well those safety processes were working or what might need to change."

Toner also recounted reports that Altman allegedly created a toxic atmosphere inside the company and allegedly behaved deceitfully toward collaborators and colleagues outside of the company.

For the board, these issues appropriately sounded alarm bells. A board that cannot trust its CEO cannot exercise its oversight duties properly, and no board should want a leader who creates a toxic environment. Perhaps most pointedly, and contrary to claims that safety issues had nothing to do with the firing, as Toner points out, a board that cannot get accurate information about safety practices has no way to effectively ensure the company was adhering to its safety-first mission.

All of these issues also speak directly to OpenAI's nonprofit status. It is almost impossible to imagine a justifiable rationale for company leadership keeping the board in the dark about a product release as explosively important as ChatGPT 3.0. On its face, and assuming Toner's claims are valid, the most likely explanation for such an action is fear that the board would have blocked or slowed release of ChatGPT 3.0, or at least that it might have done so. Such questioning from the board would have been appropriate; indeed, it was the board's duty to ask questions about, and prioritize, safety. From the management perspective, there would be one overriding rationale to rush the product release and evade potential board scrutiny: an effort to gain market share and commercial advantage.

This evaluation of Toner's disclosure is reinforced by her third, more general claim: That Altman failed to provide clear information about OpenAI's safety practices, making it impossible for the board to ensure the nonprofit mission prioritizing safety was being honored.

Toner's second example of Altman's alleged lack of candor, that he did not disclose his ownership of the OpenAI Startup Fund, is also consequential for assessing whether OpenAI is no longer functioning as a nonprofit. As I noted in my March letter to you, Altman's ownership of OpenAI Startup Fund was first reported by Axios in February. OpenAI said the situation was temporary and treated it as a kind of oversight, and Altman stepped away from his ownership position in March.[14] Even so, I noted in March, it is very hard to reconcile the known facts about OpenAI Startup Fund with the idea that the nonprofit OpenAI remains committed to its mission. Toner's remarks raise further questions about Altman's role in OpenAI Startup Fund.

Altogether, and especially in light of the information in my previous letters and in this one, Toner's comments suggest more than severe managerial issues at OpenAI: They lend strong credence to the claim that OpenAI has subordinated its nonprofit, safety-oriented mission to for-profit objectives.

---

[14] Dan Primack, "Sam Altman no longer owns OpenAI Startup Fund," Axios, April 1, 2024. https://www.axios.com/2024/04/01/sam-altman-openai-startup-fund

4. Questions About Sam Altman's Business Arrangements

OpenAI CEO and OpenAI nonprofit board member Sam Altman is paid $65,000 annually and has stated that he does not own a stake in OpenAI. However, he has an ownership stake in hundreds of tech companies and start-ups, valued in total at $2.8 billion or more.

"A growing number of Altman's startups do business with OpenAI itself, either as customers or major business partners," the Wall Street Journal reports. "The arrangement puts Altman on both sides of deals, creating a mounting list of potential conflicts in which he could personally benefit from OpenAI's work."[15]

One prospective arrangement meriting investigation is the relationship between OpenAI and Helion, a nuclear fusion start-up chaired by Altman. According to the Wall Street Journal, Altman invested $375 million in Helion in 2021. Helion's website lists Altman as one of its major investors.[16] It also states that it has raised $577 million, suggesting that Altman may even be a majority holder in the company.[17] According to the Journal, "OpenAI is in talks for a deal with Helion, a nuclear-energy startup that is chaired by Altman, in which it would buy vast quantities of electricity to provide power for data centers." The Journal also reports that Altman has recused himself from these conversations.[18]

Also worthy of investigation is the OpenAI-Reddit partnership.[19] After this partnership was announced, according to the Journal, "Reddit's stock shot up 10% … boosting Altman's stake by $69 million to $754 million."[20] OpenAI states that the partnership negotiations were led by the company's COO and "approved by its independent Board of Directors."[21]

Such arrangements raise governance and conflict of interest concerns for for-profit companies. But they are even more serious for a nonprofit, raising questions about prohibited private inurement. Too little is known about Altman's investments and the relationship of OpenAI with companies in which he owns a substantial stake. But enough is known to say the issue merits investigation, and also that it adds still more weight to the idea that OpenAI is not adhering to its nonprofit obligations.

* * *

We think this additional information further solidifies the view that OpenAI is no longer operating in accordance with its nonprofit mandate. We urge you to investigate OpenAI, Inc.'s nonprofit status expeditiously.

---

[15] Berber Jin, Tom Dotan and Keach Hagey, "The Opaque Investment Empire Making OpenAI's Sam Altman Rich," The Wall Street Journal, June 3, 2024. https://www.wsj.com/tech/ai/openai-sam-altman-investments-004fc785
[16] "Our Investors," Helion, https://www.helionenergy.com/team
[17] "FAQ," Helion, https://www.helionenergy.com/faq
[18] Jin, Dotan and Hagey.
[19] "OpenAI and Reddit Partnership," May 16, 2024, https://openai.com/index/openai-and-reddit-partnership/
[20] Jin, Dotan and Hagey.
[21] "OpenAI and Reddit Partnership."

Equally, we encourage you to commence an inquiry into OpenAI nonprofit's valuation. We note ongoing reports that OpenAI is itself considering abandoning its nonprofit status and that the OpenAI board is considering granting equity shares to Altman.[22]

As you know, and as I detailed in my January letter, if either OpenAI chooses to dissolve or convert to a for-profit corporation, or if you seek involuntary dissolution on the grounds that it is failing to carry out its nonprofit mission, then California law requires that OpenAI's assets remain perpetually dedicated to charitable enterprise. Determining the value of those assets will require a careful evaluation of the control premium that the OpenAI nonprofit maintains over the OpenAI for-profit (irrespective of its ownership stake in the for-profit), as well as the complicated financial structure of the OpenAI family of companies. The bizarre saga of OpenAI Startup Fund illustrates the cultivated complexity of that financial structure and the need to investigate whether and to what extent OpenAI nonprofit should be considered to have claim on assets in OpenAI-affiliated enterprises.

OpenAI was founded with the promise that it would be a different kind of tech firm, one that elevated and prioritized safety and the public interest over commercial considerations. There are very few serious commentators who believe OpenAI adheres to that inspiring vision any longer.[23] Indeed, its large commercial competitors routinely display more safety concern than OpenAI, as OpenAI's roll out of an anthropomorphic AI assistant illustrates. The law should catch up with the on-the-ground reality and no longer treat OpenAI as a nonprofit.

Thank you for considering this matter.

Sincerely,

Robert Weissman,
President

[22] Amir Efrati and Wayne Ma, "OpenAI CEO Cements Control as He Secures Apple Deal," The Information, May 29, 2024. https://www.theinformation.com/articles/openai-ceo-cements-control-as-he-secures-apple-deal

[23] Reports Business Insider: "[T]he startup's commercial aspirations are clear. It's aggressively pushed out new models to compete with rivals … The result, the VC told me, is people feel OpenAI is talking out of both sides of its mouth. In reality, they said, the split between OpenAI's focus on commercialization versus safety feels like it's more 95/5, respectively." Dan DeFrancesco, "Insider Today: Up In Arms Over OpenAI," Business Insider, June 5, 2024, https://www.businessinsider.com/openai-open-letter-employees-controversy-demands-sam-altman-tech-chaos-2024-06