

UPDATE
Not For Private Gain
Sept. 2026

Introduction

When the Attorneys General of California and Delaware permitted OpenAI to restructure from a nonprofit-controlled LLC to a Public Benefit Corporation (PBC), they secured 20 concessions from the company, set out in a Memorandum of Understanding with the California Attorney General (the “CA MOU”) and a Statement of No Objection from the Delaware Attorney General (the “DE Statement”)—together, the “restructuring agreements.”¹ [As we wrote at the time](#), the most important of these concessions was the obligation for directors of the new corporation to consider *only* OpenAI’s mission—ensuring artificial general intelligence (AGI) benefits all of humanity—when making safety and security decisions.² This concession is crucial in part because of what directors *cannot* consider when making such decisions: namely, the pecuniary interests of stockholders. Furthermore, OpenAI’s Safety and Security Committee—a board committee of the nonprofit with a sole duty to the mission—was tasked with “overseeing and reviewing the safety and security processes and practices of the Corporation and its controlled affiliates with respect to model development and deployment.”³ To accomplish this, the Safety and Security Committee supposedly has an effective approval right over the PBC’s actions, including halting the release of AI models.⁴ In our view, the restructuring was a major loss for the public—but the nonprofit kept this critical responsibility.

Nine months later, this governance regime faced its most significant test yet. On July 21, OpenAI disclosed that a combination of its models had escaped a sandboxed environment using a previously unknown cyber vulnerability. The models went on to chain together multiple sophisticated cyber vulnerabilities to compromise at least one other company’s infrastructure.⁵

The problem began two months prior, when models OpenAI was training discovered how to exploit their training environment to gain unauthorized access to the internet.⁶ Things got progressively worse from there. A team at OpenAI noticed agents from different sandboxes working on independent tasks communicating with each other on an improvised message board as early as May, but the significance wasn’t understood at the time. At one point, sustained agent activity crashed the compromised service, and OpenAI rebuilt it, wiping the message

¹ Cal. Dep’t of Just., [Memorandum of Understanding Between OpenAI, Inc. and the California Attorney General re Notice of Conditions of Non-Objection](#) (Oct. 27, 2025) (“CA MOU”); Del. Dep’t of Just., [Statement of No Objection of the Delaware Attorney General to OpenAI, Inc.’s Corporate Restructuring](#) (Oct. 28, 2025) (“DE Statement”).

² CA MOU ¶ 8; DE Statement ¶ 6.

³ CA MOU ¶ 9; DE Statement ¶ 7.

⁴ CA MOU ¶¶ 9, 11; DE Statement ¶¶ 7, 9.

⁵ *OpenAI and Hugging Face Partner to Address Security Incident During Model Evaluation*, OpenAI (July 21, 2026), <https://openai.com/index/hugging-face-model-evaluation-security-incident/>.

⁶ Black Hat, *The OpenAI-Hugging Face Incident* (YouTube, Aug. 6, 2026), <https://www.youtube.com/watch?t=710&v=87DyyMV0kCY&feature=youtu.be> (“Black Hat Presentation”).

board as a side effect.⁷ The agents then executed a chain of attacks allowing them to regain unauthorized access to the internet, eventually infiltrating the AI platform Hugging Face and executing over 17,000 individual actions in a weekend-long cyberattack that Hugging Face reported to law enforcement.⁸

OpenAI employees made many safety and security decisions at each step of this sequence: how to test an unreleased model of unknown capability, whether and how to supervise it, what to do upon discovering that models were coordinating to escape their own confinement, and how quickly to resume training and evaluation of the same models upon partial remediation. The Attorneys General secured commitments from OpenAI regarding precisely these types of decisions—indeed, this is the narrow domain in which total nonprofit control survived the restructuring. The nonprofit ultimately must answer for how it handled this novel and dangerous incident.

In the first section of this report, we assess what is publicly known about OpenAI’s compliance with the restructuring agreements thus far, as well as the larger mission-protection goals these agreements were designed to secure. In particular, while the nonprofit has made progress (if perhaps slowly) toward its philanthropic goals, early warning signs suggest that it is not exercising the level of oversight over the PBC that the restructuring agreements were designed to ensure, and that it may lack the staff and access necessary to do so. However, much relevant information here is not publicly available, highlighting a need for greater transparency.

To this end, one requirement of the restructuring agreements was that the OpenAI PBC publish annual public reports to demonstrate its progress toward the mission.⁹ Given that the restructuring was completed in late October 2025, OpenAI is due to publish its first report soon. In the second section of this report, we lay out what OpenAI should include in this forthcoming report so that the public and the Attorneys General can better assess its progress.

Section I: OpenAI’s adherence to its mission

Among the fundamental goals of the restructuring agreements was that the PBC would remain aligned with the mission of ensuring artificial general intelligence benefits all of humanity; the nonprofit is to provide oversight of the PBC to that end. This arrangement is reflected in a number of the provisions of the restructuring agreements, including the nonprofit’s sole power to appoint and power to remove PBC directors; the requirement that the PBC, when making safety and security decisions, consider only the mission; and the nonprofit’s oversight over these decisions via its Safety and Security Committee (SSC), including its ability to require risk mitigations up to and including halting model releases.¹⁰

⁷ Black Hat Presentation at 16:00, <https://www.youtube.com/watch?v=87DyyMV0kCY&t=960s>.

⁸ *Security Incident Disclosure – July 2026*, Hugging Face (July 16, 2026), <https://huggingface.co/blog/security-incident-july-2026>; *OpenAI and Hugging Face Partner to Address Security Incident During Model Evaluation*, OpenAI (July 21, 2026), <https://openai.com/index/hugging-face-model-evaluation-security-incident>.

⁹ CA MOU ¶ 17; DE Statement ¶ 15.

¹⁰ CA MOU ¶¶ 2, 8, 9, 11; DE Statement ¶¶ 1, 6, 7, 9.

Nearly a year after the restructuring, we have some limited evidence that the nonprofit is attempting to fulfill its designated oversight function. In particular, testimony from board member and SSC chair Zico Kolter during the *Musk v. Altman* trial revealed that the SSC had exercised its authority to delay model releases. As Kolter put it, “there were two times where we sort of formally requested a delay of models . . . in many other times as well, we have inquired about information or asked for more data prior to final review that wound up actually causing a delay of the model as well.”¹¹ Kolter also stated that there had been no pushback when the SSC exercised its authority.¹² It is encouraging that the SSC has been able to exercise these powers, though we lack details about when and why they were used.

Despite this, there are also warning signs that the nonprofit may not be exercising the level of independent oversight over the PBC that the restructuring agreements were designed to ensure and that may be necessary to prevent serious harm from OpenAI’s models.

- 1. The events surrounding the Hugging Face hack.** As previously discussed, on July 21, OpenAI revealed that a combination of its models had hacked into and compromised the infrastructure of Hugging Face, an open-source AI platform, in what OpenAI itself has called a potential loss-of-control “warning shot.”¹³ Roughly 1,200 agents that were meant to be isolated from one another found a way to communicate through an unsanctioned channel where they exchanged over 70,000 messages and files. Around 700 went on to take part in an attack on Hugging Face’s infrastructure, where they ran code on Hugging Face’s production servers, stole the company’s internal credentials, and copied several of its private code repositories. Agents also compromised OpenAI’s own research infrastructure, gaining administrator access to internal systems and reading hundreds of the company’s stored credentials.¹⁴ A week went by before OpenAI realized that its models were behind the hack, and even then, only after Hugging Face publicly disclosed the attack and reported it to law enforcement.¹⁵ OpenAI researchers subsequently revealed that the company’s AI agents had been communicating on an internal message board since May 12. The message board was not erased until early July, only for the AI agents to discover an alternative means of communication three days later.¹⁶

¹¹ [Trial Transcript](#) vol. 11, 2169–70, *Musk v. Altman*, No. 4:24-cv-04722-YGR (N.D. Cal. May 12, 2026).

¹² *Id.*

¹³ *The Hugging Face Incident and the Road Ahead*, OpenAI (Aug. 26, 2026), <https://openai.com/index/hugging-face-incident-and-the-road-ahead/>.

¹⁴ *OpenAI–Hugging Face Incident Technical Report*, OpenAI (Aug. 26, 2026), <https://cdn.openai.com/pdf/67869394-cb91-4c12-888c-5cbd85c7814c/OpenAI-Hugging-Face%20Incident-Technical-Report.pdf>; *Brief Independent Investigation of Agents’ Behavior, Reasoning and Collaboration in the OpenAI / Hugging Face Hacking Incident*, METR & Redwood Research (Aug. 26, 2026), <https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/>.

¹⁵ Raphael Satter et al., *Its AI Agent Spent Days Hacking a Company, but Sources Say OpenAI Did Not Notice for a Week*, Reuters (July 24, 2026, updated July 25, 2026), <https://www.reuters.com/business/its-ai-agent-spent-days-hacking-company-sources-say-openai-did-not-notice-week-2026-07-24/>.

¹⁶ Dana Nickel & John Sakellariadis, *OpenAI’s Models Shared Hacking Tips on a Secret Messaging Board Before Hugging Face Breach*, Politico (Aug. 5, 2026),

These types of incidents are not unique to OpenAI, as Anthropic and Meta models are now also known to have gained unauthorized access to other organizations' infrastructures after accidentally being left in an unsandboxed environment.¹⁷ But OpenAI's incident was arguably the most serious in scope, sophistication, and impact.¹⁸

The Hugging Face hack raises questions as to whether OpenAI PBC is adhering to the requirement set out in the restructuring agreements to consider only the mission in safety and security decisions and whether the SSC is exercising robust oversight over these decisions. Details released by OpenAI strongly suggest that the hack was foreseeable and avoidable. OpenAI's responders linked suspicious agent activity to the evaluation on June 27 but advised that stopping it "was not required."¹⁹ After the agents crashed an internal server on July 4, the company patched it and then resumed cyber evaluations on July 7.²⁰ One OpenAI employee admitted that "related incidents have been happening for a while."²¹ OpenAI's head of safety systems announced his departure from the company days before the hack.²² Despite prior statements emphasizing the importance of monitoring internally deployed AI systems,²³ OpenAI has acknowledged that its chain-of-thought monitors "did not run on the evaluations in this incident."²⁴ OpenAI's forthcoming model, Astra, is also rumored to use a new technique that makes the model smarter and more cost-efficient, but at the cost of undermining its

<https://www.politico.com/news/2026/08/05/openai-models-shared-hacking-tips-secret-messaging-board-hugging-face-breach-01026750>.

¹⁷ *Investigating Three Real-World Incidents in Our Cybersecurity Evaluations*, Anthropic (July 30, 2026), <https://www.anthropic.com/news/investigating-incidents-cybersecurity-evals>; Rajveer Pardesi & Mrinmay Dey, *Meta AI Model Hacks Another Company During Testing*, Reuters (Aug. 5, 2026), <https://www.reuters.com/technology/metas-ai-model-hacked-another-company-during-testing-information-reports-2026-08-05/>.

¹⁸ For example, OpenAI's incident involved finding and exploiting zero-day exploits against OpenAI's own sandboxing infrastructure, while Meta's and Anthropic's incidents stemmed from misconfigurations attributable to a third-party vendor. OpenAI's security incident was the only one to involve extensive cross-instance coordination between many separate agents exchanging tens of thousands of messages and work assignments. It was also the only incident which didn't just occur during evaluations but actually contaminated the training process itself, risking reinforcing bad behavior into a model that is in development. Complicating things, it was the only incident which was discovered at an early stage, although it was then improperly mitigated, leading to the ultimate attack.

¹⁹ *OpenAI–Hugging Face Incident Technical Report*, OpenAI (Aug. 26, 2026), <https://cdn.openai.com/pdf/67869394-cb91-4c12-888c-5cbd85c7814c/OpenAI-Hugging-Face%20Incident-Technical-Report.pdf>.

²⁰ *The Hugging Face Incident and the Road Ahead*, OpenAI (Aug. 26, 2026), <https://openai.com/index/hugging-face-incident-and-the-road-ahead/>.

²¹ Harry Booth, *How OpenAI Lost Control of an AI Model—and What Needs to Change*, Time (July 24, 2026), <https://time.com/article/2026/07/24/openai-hugging-face-attack/>.

²² Maxwell Zeff, *OpenAI's Head of Safety Is Leaving the Company*, Wired (July 10, 2026), <https://www.wired.com/story/openai-head-of-safety-leaving/>.

²³ *How We Monitor Internal Coding Agents for Misalignment*, OpenAI (Mar. 19, 2026), <https://openai.com/index/how-we-monitor-internal-coding-agents-misalignment/>.

²⁴ *The Hugging Face Incident and the Road Ahead*, OpenAI (Aug. 26, 2026), <https://openai.com/index/hugging-face-incident-and-the-road-ahead/>.

monitorability.²⁵ If true, OpenAI's decision to accept this tradeoff would be hard to square with the Hugging Face incident, which seems to have been enabled by failures in monitoring AI behavior.

We do not know whether the SSC played any role in the sequence of events that led to this failure, including whether the SSC was informed of OpenAI's prior sandboxing failures. OpenAI has said that the SSC will be providing oversight over its investigation into the Hugging Face incident, as is appropriate.²⁶ However, as we discuss below, it is unclear that the committee has the necessary resources and capacity to provide the robust oversight that these incidents clearly require.

2. **OpenAI's compliance with its Preparedness Framework is inconsistent.** The Preparedness Framework (the "PF") is OpenAI's strategy for deciding when a model is safe to release, and it has remained in effect over the past year.²⁷ For example, when OpenAI's upcoming model Astra approached critical capability in cyber operations, the company paused some of its training and evaluation workloads.²⁸

But in other respects, OpenAI has walked back its Preparedness Framework commitments. Most notably, in May 2026 OpenAI published its "Frontier Governance Framework" (the "FGF"), a watered-down version of the Preparedness Framework.²⁹ The Frontier Governance Framework is meant specifically to comply with California's SB 53 and the EU AI Act's Code of Practice.³⁰ Failure to comply with its own Frontier Governance Framework exposes OpenAI to penalties under SB 53.³¹ That the Frontier Governance Framework is a narrower, weaker version of the more robust Preparedness Framework casts serious doubt on OpenAI's commitment to those policies and practices present in its Preparedness Framework but absent in the Frontier Governance Framework, including:

²⁵ Amir Efrati et al., *OpenAI Technique in 'Astra' Model Sparks Security Concerns*, The Information (Sep. 1, 2026),

<https://www.theinformation.com/articles/secret-technique-behind-openais-astra-model-sparks-security-concerns>.

²⁶ *OpenAI and Hugging Face Partner to Address Security Incident During Model Evaluation*, OpenAI (July 21, 2026), <https://openai.com/index/hugging-face-model-evaluation-security-incident>.

²⁷ *OpenAI Preparedness Framework v2* (updated Apr. 15, 2025) ("OpenAI Preparedness Framework").

²⁸ *Pacing Model Development in an Era of Cyber-Critical Capabilities*, OpenAI (Aug. 18, 2026), <https://openai.com/index/pacing-model-development-cyber-capabilities/> ("Today, we require the strictest level of security safeguards for workloads involving Astra or cyber models, since we have determined that Astra models may have a critical level of cyber capability. . . . While some Astra training and evaluations meet those requirements, a significant number of workloads remain paused until they are fully migrated and enhanced to meet the new security bar.").

²⁹ *OpenAI's Frontier Governance Framework*, OpenAI (May 28, 2026), <https://openai.com/index/openai-frontier-governance-framework/>.

³⁰ *Id.*

³¹ *Cal. Bus. & Prof. Code § 22757.12(a)*; *id.* *§ 22757.15(a)–(b)* (authorizing civil penalties of up to \$1,000,000 per violation, including where a developer "fails to comply with its own frontier AI framework").

- Stopping further development once a model is assessed to reach critical capabilities in bio/chemical research, cyber, or AI research (PF § 2.2)
- Working with third parties to conduct evaluations and safeguard stress-testing (PF § 5.2)—the Frontier Governance Framework says only that OpenAI “*may* solicit and obtain input” (FGF § 5, emphasis added)
- Assigning the Safety Advisory Group responsibility for determining threat models, capability threshold crossings, and safeguard sufficiency (PF §§ 2.2, 3.3, 4.2)
- The Preparedness Framework applies to internally deployed models (PF § 4.4), whereas the Frontier Governance Framework largely applies to models deployed externally (FGF at 2)

The Preparedness Framework itself is outdated. OpenAI has not updated the framework since April 2025, before its restructuring. The framework says that OpenAI will review and potentially update the document “for continued sufficiency at least once a year,” yet the 17-month-old document does not reflect the basic facts of OpenAI’s corporate structure, including that there are now two OpenAI entities, the PBC and the nonprofit, and the SSC is a committee of the nonprofit’s board.³² It also does not reflect the provision negotiated by the California and Delaware Attorneys General that the SSC can require mitigatory measures, including stopping model deployment.³³

More concerning is that the Preparedness Framework is intended to be “further update[d]” before any OpenAI model reaches a critical capability level, justifying its failure to articulate any safeguards or security controls for such critical-capability models.³⁴ OpenAI has now designated Astra as having critical cyber capabilities, without having updated the framework to articulate applicable safeguards and security controls.³⁵ It also announced that it had restarted its largest frontier training run on August 28 and plans to release Astra “soon.”³⁶ The mitigatory measures OpenAI is taking for Astra—“isolated testing environments, restricted network and tool access, enhanced model weight protections and encryption, additional monitoring and detection capabilities, and sandboxed execution”—are being improvised at the moment of need rather than drawn from pre-existing standards, as the Preparedness Framework envisions.³⁷

To its credit, OpenAI seems to have realized this problem. In an update published August 18, the company announced that, in light of the Hugging Face incident and the preliminary determination that Astra may have critical cyber capabilities, OpenAI would adopt new safeguards and update the Preparedness Framework accordingly, while

³² [OpenAI Preparedness Framework](#), Appendix B.

³³ CA MOU ¶ 11; DE Statement ¶ 9.

³⁴ [OpenAI Preparedness Framework](#), § 4.4.

³⁵ *Path to Astra: Critical Capabilities and Frontier Safeguards*, OpenAI (Sep. 1, 2026), <https://openai.com/index/path-to-astra/>.

³⁶ *Id.*

³⁷ *Responding to the Next Frontier of Critical Cyber Capabilities*, OpenAI (Aug. 7, 2026), <https://openai.com/index/responding-next-frontier-critical-cyber-capabilities/>.

taking steps to slow its model training in the meantime.³⁸ But that sequencing is itself a problem. The fact that the Preparedness Framework is being updated in response to a serious incident, rather than before one, suggests it is not serving its intended purpose.

OpenAI's compliance with the Preparedness Framework has also been called into question by the release of GPT-5.3-Codex, which some have claimed violates California's AI safety laws.³⁹ There has also been significant turnover among OpenAI's safety talent.⁴⁰

- 3. The nonprofit appears to lack dedicated staff for PBC oversight.** In our last update, we identified one key indicator of whether OpenAI was taking the steps necessary to advance its charitable mission: whether it provides staffing and resources to the nonprofit to oversee the PBC, particularly to the SSC.⁴¹ This need is also reflected in OpenAI's restructuring agreements, which required that the PBC provide the nonprofit with the resources and personnel reasonably required to advance its mission.⁴² Given that the SSC is composed of only part-time board members, having additional personnel to support its work is particularly important. However, there is no public evidence that the nonprofit has any dedicated personnel to support its oversight capacity, despite having made a number of hires to support its philanthropic grantmaking activities and general operations.⁴³ The nonprofit also has yet to appoint a CEO, which is likely inhibiting its oversight capacity.⁴⁴
- 4. OpenAI has yet to designate a second nonprofit-only board director.** As we noted in our last update, one of the nonprofit's most significant challenges in providing independent oversight of the PBC is the total overlap between its board and the PBC's board, with the exception of Zico Kolter, who serves only on the nonprofit board.⁴⁵ In light of this, the restructuring agreements require that within a year of the restructuring, at

³⁸ *Pacing Model Development in an Era of Cyber-Critical Capabilities*, OpenAI (Aug. 18, 2026), <https://openai.com/index/pacing-model-development-cyber-capabilities/>.

³⁹ Beatrice Nolan, *OpenAI Disputes Watchdog's Claim It Violated California's New AI Safety Law With Latest Model Release*, Fortune (Feb. 10, 2026), <https://fortune.com/2026/02/10/openai-violated-californias-ai-safety-law-gpt-5-3-codex-ai-model-watchdog-claims/>.

⁴⁰ Madison Mills, *OpenAI Sheds Senior Execs in Pre-IPO Refresh*, Axios (Aug. 14, 2026), <https://www.axios.com/2026/08/14/openai-executive-greg-brockman-ipo>; Maxwell Zeff, *OpenAI's Head of Safety Is Leaving the Company*, Wired (July 10, 2026), <https://www.wired.com/story/openai-head-of-safety-leaving/>.

⁴¹ *Statement on OpenAI's Restructuring*, Not For Private Gain (Nov. 2025), <https://notforprivategain.org/november-update> ("NFPG November Update").

⁴² CA MOU ¶ 13; DE Statement ¶ 11.

⁴³ During *Musk v. Altman*, SSC chair Zico Kolter confirmed that the nonprofit had only made its first full-time hires a few months before the trial, which appears to correspond to a March announcement which named a number of such hires. (See [OpenAI's March 2026 update on the Foundation](#).) Since then, OpenAI has not publicly announced any new hires for the nonprofit.

⁴⁴ Bret Taylor, *Update on the OpenAI Foundation*, OpenAI Found. (Mar. 24, 2026), <https://openaifoundation.org/news/update-on-the-openai-foundation>.

⁴⁵ [NFPG November Update](#).

least one OpenAI board director besides Zico Kolter will sit only on the nonprofit board.⁴⁶ Ten months after the restructuring, OpenAI has yet to designate such a director, though it has appointed two new directors to both the PBC and nonprofit boards, David Vélez and Robin Vince.⁴⁷ Both of the new directors also run companies that are significant enough OpenAI customers that they are profiled on OpenAI's website, which may create conflicts of interest.⁴⁸

Section II: OpenAI's forthcoming report on mission progress

Another concession secured by the Attorneys General requires the OpenAI PBC to produce an annual public report demonstrating its progress toward OpenAI's charitable mission.⁴⁹ Because the restructuring was finalized in late October 2025, the PBC is due to produce its first report soon. Given how much we do not know about OpenAI's compliance with the agreements and its behavior in accordance with the mission, the PBC should treat this report not as a box-checking exercise but as a genuine opportunity to inform its beneficiaries: the public.

Given that nonprofit oversight is perhaps the most important mechanism OpenAI's structure provides to ensure the PBC's adherence to the mission, the PBC should include details on that oversight in this report, including the role the SSC has played.

Below are eight topics that the report should cover relevant to the PBC's adherence to the restructuring agreements and nonprofit control. We encourage the Attorneys General to take careful notice of what OpenAI discloses with respect to these topics.

1. **The scope of "safety and security" decisions.** The OpenAI PBC is required to consider only its mission "in respect of safety and security issues."⁵⁰ The SSC has an approval right "over PBC actions relating to safety and security."⁵¹ OpenAI has not, however, publicly described how it interprets the scope of "safety and security" in these contexts. We therefore do not know what approval rights the SSC has in practice. Because these rights were secured by contract, OpenAI can clarify this easily by disclosing the relevant contractual language.
2. **How and when the SSC has exercised its authority.** The SSC is entitled to "oversee[] and review[] the safety and security processes and practices" of the PBC.⁵² As detailed earlier, we still do not know much about how the SSC has exercised its rights. OpenAI

⁴⁶ CA MOU ¶ 14; DE Statement ¶ 12.

⁴⁷ *David Vélez and Robin Vince Join the Boards of the OpenAI Foundation and OpenAI Group PBC*, OpenAI (July 21, 2026), <https://openai.com/index/david-velez-robin-vince-join-openai-boards/>.

⁴⁸ Regarding these companies' partnerships with OpenAI, see OpenAI's posts [about Nubank](#) and [about BNY](#). Of note, the agreements (CA MOU ¶ 15 and DE Statement ¶ 13) give the PBC board the authority to make determinations as to whether a conflict of interest compromises director independence.

⁴⁹ CA MOU ¶ 17; DE Statement ¶ 15.

⁵⁰ CA MOU ¶ 8; DE Statement ¶ 6.

⁵¹ CA MOU ¶ 9; DE Statement ¶ 7.

⁵² *Id.*

should clarify this. In particular, OpenAI should (a) clarify the SSC's involvement in the decisions leading to the Hugging Face hack, (b) reveal whether the SSC reviewed and had input into OpenAI's contract with the Pentagon,⁵³ and (c) disclose when and how the SSC exercised its authority to require risk mitigations or to halt or delay model releases. OpenAI should also release the nonprofit board's Unanimous Written Consent detailing the scope of the SSC's authority, which would shed light on the extent to which the committee is fulfilling its defined role.⁵⁴

3. **Resources provided for SSC oversight of the PBC.** The restructuring agreements require the PBC to provide the nonprofit with the resources and personnel reasonably required for its mission.⁵⁵ As described earlier, however, it is unclear that the nonprofit (and in particular the SSC) has *any* dedicated personnel to support its oversight of the PBC. Given the severity and number of recent incidents in which OpenAI models have compromised the company's safeguards, it is unclear that an SSC composed only of part-time board members is equipped to provide oversight of the company's safety and security decisions. OpenAI should clarify what specific resources it has provided the SSC—including funding, personnel, and access to OpenAI's own technology—and if the pattern of recent incidents has led it to provide additional resources for the committee. It should provide some specific numbers to illustrate the magnitude of the support so this can be tracked over time, instead of generically describing the categories of support.
4. **How the SSC was informed of safety and security incidents.** As we described earlier, the Hugging Face breach occurred after a series of concerning safety and security incidents within OpenAI. OpenAI should disclose what steps it took to keep the SSC informed as these incidents were discovered. OpenAI should also disclose what actions, if any, the SSC took in response to these incidents.
5. **How directors are managing their dual loyalties.** As we identified in our last update, a possible impediment to effective nonprofit control over the PBC is the near-total overlap in board membership.⁵⁶ As one step towards addressing this, the restructuring agreements require that dual-board directors attend an annual training on their fiduciary duties with respect to each board.⁵⁷ The report should describe how directors manage their dual loyalties, including incidents in which this conflict has come up and the decisions that were made as a result, and details like whether the nonprofit board has its own counsel.
6. **How OpenAI would operationalize its stop-and-assist commitment.** The restructuring agreements require the PBC to adopt the OpenAI Charter as the principles

⁵³ Sharon Goldman, *OpenAI Strikes A Deal With The Pentagon Just Hours After Trump Orders The End Of Anthropic Contracts, And Hours After A Staff All-Hands*, Fortune (Feb. 27, 2026), <https://fortune.com/2026/02/27/openai-in-talks-with-pentagon-after-anthropic-blowup/>.

⁵⁴ CA MOU ¶ 9; DE Statement ¶ 7.

⁵⁵ CA MOU ¶ 13; DE Statement ¶ 11.

⁵⁶ [NFGP November Update](#).

⁵⁷ CA MOU ¶ 21; DE Statement ¶ 19.

it will use to execute its mission.⁵⁸ The Charter's stop-and-assist commitment—to stop competing with and start assisting a value-aligned, safety-conscious project that comes close to building AGI first—distinguishes OpenAI's mission from that of an ordinary AI company.⁵⁹ The report should confirm whether OpenAI remains committed to this and describe what steps OpenAI has taken to prepare to operationalize it.

7. **Whether an IPO would affect the nonprofit's control.** As per the restructuring agreements, the nonprofit board retains the sole power to appoint, and the power to remove, PBC directors.⁶⁰ OpenAI is reportedly planning an initial public offering in 2027 or sooner.⁶¹ The report should confirm that the nonprofit will retain these powers conferred by its Class N stock under the post-IPO structure.⁶² The report should also confirm that the contractual rights given to the SSC will survive the IPO.
8. **How the PBC makes director independence determinations.** The restructuring agreements require that OpenAI have robust corporate governance guidelines and conflict of interest policies, and that the PBC board have three committees (Audit & Risk, Compensation, and Nominating & Governance) composed of independent directors.⁶³ In addition, they require that a majority of the PBC board be independent directors, while giving the board the authority to decide whether an external relationship or interest compromises director independence.⁶⁴ Given the concerns regarding potential conflicts of interest among PBC board members, the report should clarify which directors are considered independent, the standard the board uses to determine director independence, and the membership of the independent-director committees.

OpenAI should also promptly name a second nonprofit-only board director, in accordance with the terms of the restructuring agreements, as well as an independent nonprofit CEO. These should be individuals with expertise not just relevant to the nonprofit's philanthropic duties but also to its PBC oversight function, including its safety and security authority.

Conclusion

In light of OpenAI's recent record of safety and security incidents, it is clear that the safeguards preserved by the Attorneys General during OpenAI's restructuring are more important than ever. While notably weaker than those in OpenAI's original structure, they nonetheless represent an important mechanism for ensuring that OpenAI remains aligned with its charitable mission.

⁵⁸ CA MOU ¶ 4; DE Statement ¶ 3.

⁵⁹ OpenAI Charter, <https://openai.com/charter/>.

⁶⁰ CA MOU ¶ 2; DE Statement ¶ 1.

⁶¹ Ashley Capoot & Kate Rooney, *OpenAI 'Will Be A Public Company In 2027' Or Sooner, CFO Friar Tells Employees*, CNBC (Aug. 19, 2026), <https://www.cnbc.com/2026/08/19/open-ai-ipo-timing-2027-friar.html>.

⁶² CA MOU ¶¶ 2, 7; DE Statement ¶¶ 1, 5.

⁶³ CA MOU ¶ 16; DE Statement ¶ 14.

⁶⁴ CA MOU ¶ 15; DE Statement ¶ 13.

OpenAI's first annual report on mission progress is an opportunity to demonstrate that it takes its obligations to carry out its charitable mission seriously, rather than a set of constraints to be managed. We trust that the Attorneys General will take notice of whether OpenAI fails to carry out these commitments.

Appendix A: OpenAI Foundation's dual activities of philanthropy and governance

While OpenAI's restructuring removed the profit cap model that limited how much influence investors would have in futures where OpenAI grows exceptionally powerful, it did have the positive benefit of endowing the nonprofit with a substantial equity stake, currently valued at close to \$200 billion, that will allow it to begin spending money sooner.⁶⁵

The body of this report concerns the OpenAI Foundation's oversight of the OpenAI Public Benefit Corporation. We believe that oversight constitutes the nonprofit's most important, most at-risk, and least visible function. But the nonprofit's philanthropic work represents the other half of its responsibilities, and it is the half that OpenAI has been most willing to discuss. Below, we set out what is publicly known about that work, including the steps we see as promising and those that give us pause.

Signs of progress:

- 1. Grantmaking areas aligned with OpenAI's mission.** The restructuring agreements require that the directors of the nonprofit have a fiduciary responsibility to the nonprofit, its mission, and its beneficiaries.⁶⁶ The grantmaking areas that the nonprofit has announced—life sciences and curing diseases, jobs and economic impact, AI resilience, and supporting communities—are aligned with this.⁶⁷
- 2. Qualified hires.** OpenAI has also announced a number of promising early hires to lead these portfolios. This includes bringing on Wojciech Zaremba, computer scientist and OpenAI cofounder, to lead its AI resilience portfolio, and Jacob Trefethen, an experienced health and science grantmaker, as Head of Life Sciences and Curing Diseases.⁶⁸

There are also a few warning signs regarding how the nonprofit has approached its philanthropic giving, and how this intersects with the interests of the PBC.

⁶⁵ As of the closing of the recapitalization in October of 2025, the OpenAI Foundation held a 26% equity stake in OpenAI Group, worth approximately [\\$130 billion](#). In March of 2026, OpenAI raised \$122 billion at a post-money valuation of [\\$852 billion](#). A 26% stake at the most recent valuation would be worth \$221.5 billion. However, if the entire \$122 billion was newly issued equity, the OpenAI Foundation's stake would be 22.3% of the company or approximately \$190 billion.

⁶⁶ CA MOU ¶ 5; DE Statement ¶ 4.

⁶⁷ Bret Taylor, *Update on the OpenAI Foundation*, OpenAI Found. (Mar. 24, 2026), <https://openaifoundation.org/news/update-on-the-openai-foundation>.

⁶⁸ *Id.*

Causes for concern:

- 1. The slow pace of the nonprofit's hiring and grantmaking.** As noted earlier, the OpenAI Foundation has yet to appoint a CEO and made its first full-time hires only in March, nearly five months after the restructuring. That same month, the nonprofit also announced it planned to spend at least \$1 billion in the next year.⁶⁹ While substantial, this amount is less than 1% of the nonprofit's roughly \$200 billion valuation and significantly less than the 5% that a private foundation would be required to deploy.⁷⁰ This is particularly surprising given the relatively short amount of time that OpenAI leadership believes we have before AGI arrives; in a recent interview, Sam Altman said that he expects the company to have an internal system that he would call AGI by the end of the year.⁷¹ One would expect that increasing the velocity of their grantmaking in this critical period would be a top priority for the organization.
- 2. The potential for spending nonprofit funds when PBC funds make more sense.** In June, the OpenAI Foundation announced that it would be a founding partner of RAISE US, a national nonprofit that will help the US workforce adapt to AI.⁷² Many of the other large tech companies that were founding partners, however, appeared to be contributing directly from their for-profit arms—including Amazon, Anthropic, and Microsoft.⁷³ To the extent that the OpenAI Foundation is spending on things that the OpenAI PBC can and would have spent on itself, it is failing to fully deliver on its mission; the nonprofit was designed to be a vehicle for the preservation of OpenAI's unique charitable purpose, not just an ordinary PR or government affairs arm. Another salient concern is regarding OpenAI's proposal that multiple AI companies give the Trump administration a 5% equity stake; OpenAI has not clarified whether this would come from issuing new shares or transferring shares, including shares currently held by the nonprofit.⁷⁴ If OpenAI's stake came from the nonprofit, this would disproportionately benefit the OpenAI PBC relative to its competitors, since its other investors would not be diluted.
- 3. OpenAI's own framing of the nonprofit as a grantmaking organization.** During the *Musk v. Altman* trial, Bret Taylor—the chair of the PBC and nonprofit boards—was asked

⁶⁹ *Id.*

⁷⁰ *Taxes on Failure to Distribute Income - Private Foundations*, IRS, <https://www.irs.gov/charities-non-profits/private-foundations/taxes-on-failure-to-distribute-income-private-foundations>.

⁷¹ Alex Heath, *Inside OpenAI's Reboot*, Time (Aug. 26, 2026), <https://time.com/article/2026/08/26/openai-sam-altman-interview/>.

⁷² Gina Raimondo and Eric Holcomb Launch RAISE US, *Uniting the Nation's Leading Employers and Bipartisan Governors Behind American Workers*, Rockefeller Found. (June 25, 2026), <https://www.rockefellerfoundation.org/news/raise-us-launches-uniting-nations-leading-employers-and-bipartisan-governors-behind-american-workers/>.

⁷³ *Id.*

⁷⁴ Siladitya Ray, *OpenAI Reportedly Pitches Granting U.S. Government 5% Stake*, Forbes (July 2, 2026), <https://www.forbes.com/sites/siladityaray/2026/07/02/openai-reportedly-pitches-granting-us-government-5-stake/>.

about whether the nonprofit had staff dedicated to AI safety research. Taylor indicated it did not, saying that “our public benefit corporation is the part of OpenAI that does AI research and ensures safety” and that the nonprofit is “not a research organization” but a “grantmaking organization.”⁷⁵ While it is true that the nonprofit is not designed to do safety research per se, it is also notable that Taylor framed the PBC as the part of OpenAI that “ensures safety” while referring to the nonprofit as merely a “grantmaking organization.” This framing ignores the commitments OpenAI made in its restructuring agreements and reflects a concerning divergence between those commitments and how the nonprofit appears to be viewed internally.

Because OpenAI’s philanthropic giving is the side of the nonprofit’s work the company has chosen to make most visible, it is likely to feature prominently in the forthcoming annual report. The report should make the case for why the spending furthers the mission—both that it is happening fast enough and also that it is being targeted toward the most important opportunities available. The nonprofit’s charitable purpose is not discharged by spending alone, however, and OpenAI’s report should also address how the nonprofit is discharging its oversight and governance responsibilities.

Appendix B: A point-by-point breakdown of OpenAI’s compliance with its restructuring agreements

Below is a point-by-point account of what is publicly known about OpenAI’s compliance with its restructuring agreements, as of the date of this report.

1. **OpenAI keeps its nonprofit and PBC headquarters in California.** (*CA MOU* ¶ 1). *Compliant.* OpenAI remains headquartered in San Francisco.
2. **The nonprofit retains control of the PBC through Class N stock, with the sole power to appoint and the power to remove PBC directors.** (*CA MOU* ¶ 2; *DE Statement* ¶ 1). *Compliant.* The nonprofit continues to hold Class N stock and retains the power to appoint and remove directors. Whether that control is exercised in practice is the subject of Section I. The report should confirm that these powers will survive the planned IPO and any future rearrangement of the share classes.
3. **The PBC’s mission is identical to the nonprofit’s, and the nonprofit’s mission is unchanged by the recapitalization.** (*CA MOU* ¶ 3; *DE Statement* ¶ 2). *Compliant.* There have been no public updates to the OpenAI mission since the restructuring. At times, the company describes its mission as building AGI, when it is really to ensure that AGI benefits humanity whether or not they build it.⁷⁶ This distinction should continue to be respected in future reports from the nonprofit.

⁷⁵ [Trial Transcript](#) vol. 11, 2004–05, *Musk v. Altman*, No. 4:24-cv-04722-YGR (N.D. Cal. May 12, 2026).

⁷⁶ *Federal Account Director, National Security*, OpenAI (archived Sept. 2, 2026), <https://web.archive.org/web/20260902055429/https://openai.com/careers/federal-account-director-national-security-washington-dc/>; Catherina Gioino, *OpenAI Changed Its Mission Statement 6 Times In 9 Years. It Finally Removed The Word “Safely” As A Core Value When It Restructured Into A For-Profit*, *Fortune* (Feb. 23, 2026), <https://fortune.com/2026/02/23/openai-mission-statement-changed-restructuring-forprofit-business/>.

4. **The PBC publishes the OpenAI Charter.** (CA MOU ¶ 4; DE Statement ¶ 3). *Compliant.* The Charter remains posted on the OpenAI website, and it still contains the stop-and-assist commitment. OpenAI has given little public indication of how it currently regards that obligation, which Section II asks it to address directly.
5. **Nonprofit directors' fiduciary duties are to the nonprofit, the mission, and the public.** (CA MOU ¶ 5; DE Statement ¶ 4). *Presumed compliant.* There have been no changes announced to the nonprofit's governance since the restructuring. Because most nonprofit directors also sit on the PBC board, the report should describe how they balance the two sets of duties, including whether the nonprofit board has retained its own counsel.
6. **The nonprofit holds charitable assets in California for the benefit of Californians.** (CA MOU ¶ 6). *Presumed compliant.* OpenAI's charitable assets are presumed to still be held in California. If OpenAI proceeds with its proposal that AI companies give the federal government a 5% equity stake, it should confirm that its own stake will not come from the nonprofit's assets.
7. **The nonprofit's written approval is required before the PBC amends its governance guidelines, changes the mission, sells or merges material assets, issues more Class N shares, creates new director-voting stock, or amends its charter to the detriment of the Class N holder.** (CA MOU ¶ 7; DE Statement ¶ 5). *Presumed compliant.* There are no public signs that any of the six actions have occurred without approval. Should the IPO involve any of these actions, the nonprofit should publish its written approval of those actions.
8. **The PBC charter requires directors to consider only the mission, and not the pecuniary interests of stockholders, on safety and security matters.** (CA MOU ¶ 8; DE Statement ¶ 6). *Potentially noncompliant.* This provision exists in the charter, but various incidents at the company over the past year, including the Hugging Face incident, call into question whether the PBC is adequately prioritizing safety and security in practice.
9. **The SSC is a committee of the nonprofit with safety and security oversight and a contractual approval right over PBC actions.** (CA MOU ¶ 9; DE Statement ¶ 7). *Presumed compliant.* The SSC remains a committee of the nonprofit board. The restructuring agreements with the Attorneys General provided that it would have a contractual approval right, but OpenAI has not revealed anything about that contractual right, including whether it remains in place.
10. **The SSC chair sits only on the nonprofit board and has full observation rights at PBC meetings.** (CA MOU ¶ 10; DE Statement ¶ 8). *Presumed compliant.* Zico Kolter remains in this role and presumably has observation rights.
11. **The SSC may require risk mitigations, including halting model releases.** (CA MOU ¶ 11; DE Statement ¶ 9). *Compliant.* Kolter testified in *Musk v. Altman* that the committee had formally requested delays twice, and had caused others by asking for more information before final review.⁷⁷
12. **OpenAI continues to mitigate risks to teens and others.** (CA MOU ¶ 12; DE Statement ¶ 10). *Partially compliant.* OpenAI has taken a number of steps to reduce

⁷⁷ [Trial Transcript](#) vol. 11, 2169–70, *Musk v. Altman*, No. 4:24-cv-04722-YGR (N.D. Cal. May 12, 2026).

risks to minors.⁷⁸ It has also faced litigation alleging that those measures are inadequate, including from the Florida Attorney General,⁷⁹ and experts have raised similar concerns.⁸⁰

13. **The PBC must provide the nonprofit with personnel and resources at no cost, under a Support and Services Agreement.** (CA MOU ¶ 13; DE Statement ¶ 11). *Potentially noncompliant.* The agreement is not public, and little is known about what the PBC has provided under it. What is visible points to a nonprofit under-resourced for the job, without dedicated staff supporting oversight of the PBC or the SSC, and no chief executive officer. Section I further explains our concerns; Section II asks OpenAI to give more detail about the type and magnitude of the support it has provided.
14. **Within one year of the restructuring, at least one nonprofit director besides the SSC chair will serve only on the nonprofit board.** (CA MOU ¶ 14; DE Statement ¶ 12). *To be determined.* The deadline falls in late October, leaving OpenAI a bit more time to comply. No such director has been designated, and the two directors added in July serve on both boards.
15. **The PBC board will be composed of a majority of independent directors.** (CA MOU ¶ 15; DE Statement ¶ 13). *Potentially noncompliant.* Sam Altman is the only board member who is an employee or member of management, so by a restrictive definition of independence, the provision's requirement is met. However, a majority of the current nonprofit and PBC board members are entangled with OpenAI's business—including running companies that are significant customers or holding interests in the sector—to a degree that could bear on their judgment when the mission and the company's pecuniary interests diverge.⁸¹
16. **OpenAI will maintain governance guidelines and conflict of interest policies, and establish Audit & Risk, Compensation, and Nominating & Governance committees composed solely of independent directors.** (CA MOU ¶ 16; DE Statement ¶ 14).

⁷⁸ *Introducing ChatGPT for Teens: Built for Learning, Backed By Protections*, OpenAI (Aug. 18, 2026), <https://openai.com/index/chatgpt-for-teens/>; *Introducing Parental Controls*, OpenAI (Sep. 29, 2025), <https://openai.com/index/introducing-parental-controls/>.

⁷⁹ *Attorney General James Uthmeier Files First-in-the-Nation State-Led Lawsuit Against OpenAI, CEO Sam Altman for Deceptive Practices and Harms to Floridians*, Fla. Off. of the Att'y Gen. (June 1, 2026), <https://www.myfloridalegal.com/newsrelease/attorney-general-james-uthmeier-files-first-nation-state-led-lawsuit-against-openai-ceo>.

⁸⁰ *Common Sense Media Report Finds ChatGPT and Sora Pose Risks to Teens Despite Safety Features*, Common Sense Media (Oct. 23, 2025), <https://www.common Sense Media.org/press-releases/common-sense-media-report-finds-chatgpt-and-sora-pose-risks-to-teens-despite-safety-features>; Igor Bonifacic, *Child Safety Experts Are Skeptical of OpenAI's ChatGPT for Teens*, Engadget (Aug. 21, 2026), <https://www.engadget.com/2241179/child-safety-experts-skeptical-of-openai-chatgpt-for-teens/>.

⁸¹ The ten current directors of the OpenAI Foundation are Bret Taylor, Sam Altman, Adam D'Angelo, Sue Desmond-Hellmann, Zico Kolter, Paul Nakasone, Adebayo Ogunlesi, Nicole Seligman, David Vélez, and Robin Vince; all but Kolter also serve on the PBC board. Six of the ten have commercial ties to OpenAI or its market: Altman holds investments in OpenAI partners and suppliers including Helion Energy and Retro Biosciences; Taylor is CEO of Sierra, which builds chatbots using OpenAI's models; D'Angelo is CEO of Quora, whose Poe platform is a major OpenAI customer; Ogunlesi's Global Infrastructure Partners is a partner in a \$30 billion AI data-center fund; and Vélez and Vince lead [Nubank](#) and [BNY](#), both OpenAI enterprise customers. For documentation of the remaining ties, see the [CEO & Board Conflicts of Interest section of The OpenAI Files](#), published by The Midas Project & The Tech Oversight Project.

Compliance uncertain. There is no public confirmation that the three committees exist or who sits on them, and OpenAI has not disclosed the details of its conflict of interest policy.

17. **The PBC will publish fully public reports, at least annually, demonstrating progress towards the mission.** (*CA MOU ¶ 17; DE Statement ¶ 15*). *To be determined.* No report has been published. This is the subject of Section II.
18. **Nonprofit and PBC leadership will meet with the Attorneys General's office regularly.** (*CA MOU ¶ 18; DE Statement ¶ 16*). *Presumed compliant.* Nothing about the cadence or content of these meetings is public.
19. **The nonprofit will give 21 days' notice before consenting to a change of control, a change to the PBC mission, or an amendment reducing the Class N rights—and, under the California agreement, before relocating either headquarters out of California.** (*CA MOU ¶ 19; DE Statement ¶ 17*). *Presumed compliant.* No action triggering the requirement is known to have occurred, and no amendments to the PBC charter have occurred since the restructuring.
20. **Directors serving on both boards must complete annual fiduciary-duty training, delivered by the PBC's Chief Compliance Officer and outside counsel.** (*CA MOU ¶ 21; DE Statement ¶ 19*). *Presumed compliant.* There is no public confirmation that the training has occurred, and none would be expected. Section II asks OpenAI to confirm it in the report and to describe how dual-board directors manage their competing duties.

The remaining provisions not discussed above are standard contractual provisions, as well as one entitling the Attorney General to retain experts at the nonprofit's expense.⁸²

⁸² CA MOU ¶ 20; DE Statement ¶ 18.