AINOW

Public Good or Private Gain?

Corporate Structure and Mission Accountability at Leading AI Companies

By Maroussia Lévesque

Assistant Professor of Law & AI, Queen's University

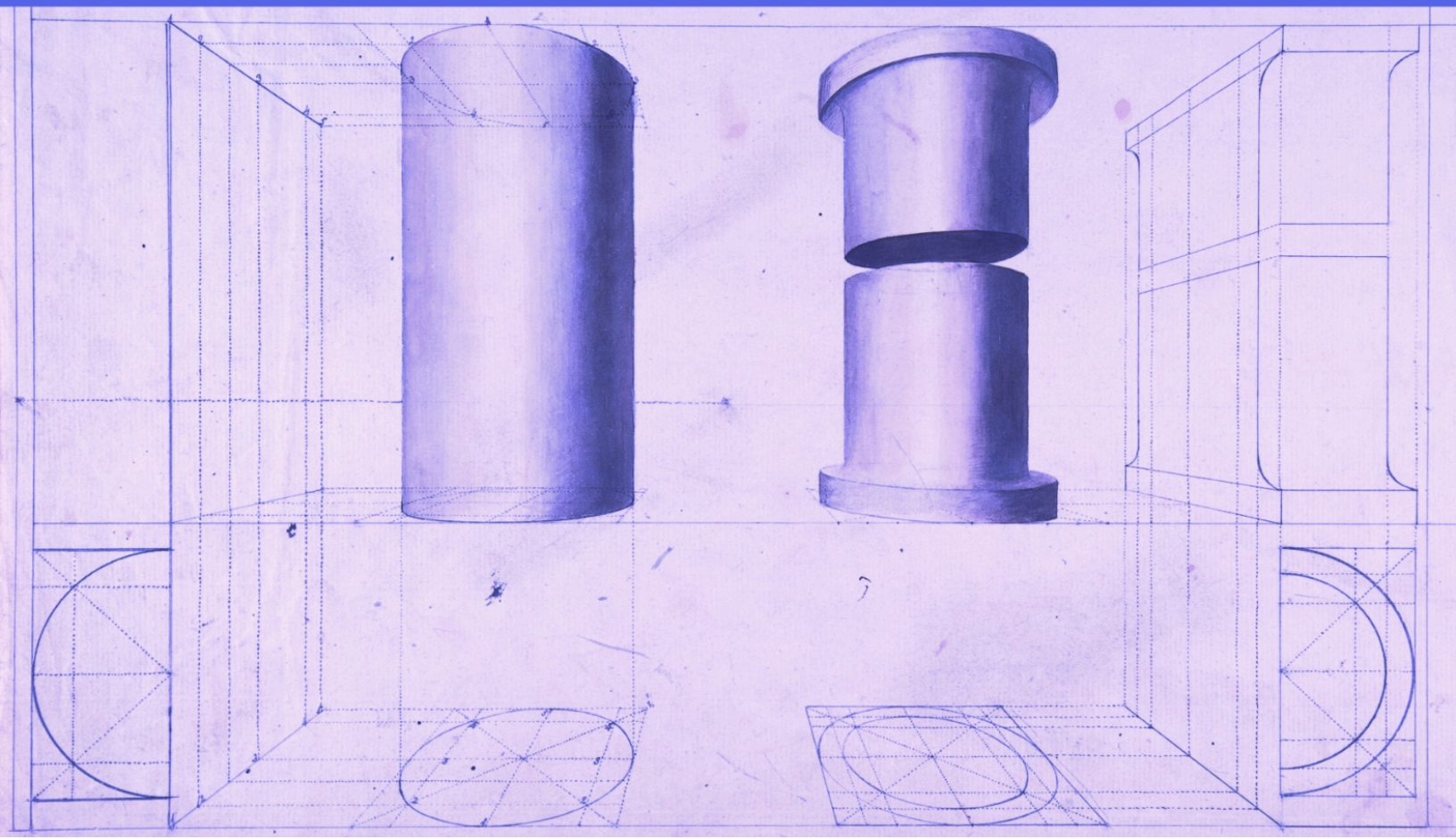


Table of Contents

Introduction	1
Part I. Taking Stock of Business Structures	5
Overview	8
DeepMind/Google	12
OpenAI	22
Anthropic	28
Part II. Business Structures in Action: Assessing Corporate Practices	
AI Regulation	
Voluntary Scaling Policies	
Military AI	
Part III. Levers to Enforce the Mission	
General Shareholder Recourse and Mandatory Disclosures	
Protecting the Charitable Mission of a Non-Profit	
Enforcing the Mission of a Public Benefit Corporation	
Revoking Corporate Charters	
Conclusion	
Annex 1: Google/DeepMind — Structure & Conduct Timeline	
Annex 2: OpenAI — Structure & Conduct Timeline	
Annex 3: Anthropic — Structure & Conduct Timeline	
Credits	

Introduction

> "Our mission is to build AI responsibly to benefit humanity"[1] — Google DeepMind
>
> "Our primary fiduciary duty is to humanity"[2] — OpenAI
>
> "The specific public benefit that the Corporation will promote is to responsibly develop and maintain advanced AI for the long-term benefit of humanity"[3] — Anthropic

Leading AI companies all purport to "benefit humanity." Though each company frames its mission slightly differently,[4] at a high level they signal remarkably similar public-minded aspirations. This report addresses whether the distinct business structures of major AI companies have any bearing on their concrete actions. The legal framework under which these U.S.-based AI businesses operate varies among them and over time, offering an opportunity to compare how different business structures align mission with action. Google/Alphabet is a traditional for-profit corporation; OpenAI morphed from a non-profit to a hybrid 'capped-profit' and recently to a public benefit corporation seeking to combine profit and social objectives, whereas Anthropic combines a public benefit corporation with a long-term benefit trust. Yet there is one striking commonality across leading AI companies: they all proceed from the thesis that the broad societal impact of AI is too important to be left to purely corporate incentives. At the end of the day, however, they all ultimately move to the beat of the market, resolving any tensions in favor of financial interest.

AI companies diverge as to how they structure their activities to achieve their mission, but ultimately converge towards for-profit goals. AI being a capital-intensive industry, companies need to fundraise to sustain the ever-rising costs of training and running AI systems.[5] Notwithstanding the shaky

[1] *About Google DeepMind*, Google DeepMind, https://deepmind.google/about/ (last visited July 3, 2025).
[2] *OpenAI Charter*, https://openai.com/charter/ (last visited June 23, 2025).
[3] Anthropic Certificates of Incorporation, https://drive.google.com/drive/folders/1ImqXYv9_H2FTNAujZfu3EPtYFD4xlIHJ, [https://perma.cc/9TGX-PRGZ] (last visited July 8, 2025).
[4] About Google DeepMind, *supra* note 1 ("OUR MISSION Build AI responsibly to benefit humanity"); *see also* Sundar Pichai, *AI at Google: Our Principles*, Google (June 7, 2018), https://www.blog.google/technology/ai/ai-principles/, [https://perma.cc/3ALF-WTEC]; for the full description of OpenAI's mission, see OpenAI, *Internal Revenue Service Form 1023: Application for Recognition of Exemption Under Section 501(c)(3) of the Internal Revenue Code* 31 (2016), https://www.documentcloud.org/documents/25197523-openai-application-for-tax-exempt-status/ (stating that OpenAI's "goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole"); *see also OpenAI Charter, supra* note 2 (describing OpenAI's goal using identical language); Anthropic Certificates of Incorporation, *supra* note 3 (describing Anthropic's proposed "specific public benefit" as being "to responsibly develop and maintain advanced AI for the long-term benefit of humanity").
[5] *Deals - Anthropic - Company Profile*, https://my.pitchbook.com/profile/466959-97/company/deals#deal-287700-58T (last visited Sep. 11, 2025) (showing that Anthropic raised over $30 billion to date, with a $13 billion funding round in September 2025).

fundamentals, these companies still achieve astronomical valuations and investor interest.[6] In light of this financial pressure, it's unclear whether any of them have the right incentives to prioritize aspirational goals to "benefit humanity" when that would adversely impact their bottom line. The 244-billion-dollar question—the estimated value of the current AI market[7]—is: do AI companies walk the talk when purpose collides with profit? How does their business structure constrain or enable them to do so?

Overall, the report finds that business structure alone is insufficient to ensure companies abide by their mission. Structure does create friction, but capital eventually overcomes it. OpenAI's transformation speaks to this dynamic. The attorneys general leveraged their jurisdiction over non-profits, citing their ability to block the company's restructuring to secure commitments to the mission. Be that as it may, the original structure was dismantled precisely because it worked. When the non-profit became an impediment to raising capital, the company conceded some mission safeguards as a compromise to break away from its capital-raising constraints. In fact, financial considerations deeply shape the business structure of all AI companies under review. Through a series of case studies, this research finds that as AI development goes from proof of concept to product, AI companies consistently modify their business structures so as to become more responsive to financial considerations and reduce mission constraints on their actions. This creates deep tension within the market structure for AI: at the moment where these companies are most poised to enact their mission, they evolve their corporate governance to free themselves from any commitment to achieving it.

This suggests that the ability to evolve out of a constraining corporate form is itself a structural defect of our corporate governance system. Private, reversible choices as to the legal framework companies operate under are too brittle to operationalize public missions. Though various stakeholders have mobilized the levers of corporate law to secure some limits on the behavior of AI companies, especially when it comes to attorneys general using their ability to enforce the non-profit form of OpenAI to secure some safety commitments from the company, these efforts remain few and far between. Without binding, capture-proof regulation, corporate mission statements ring hollow.

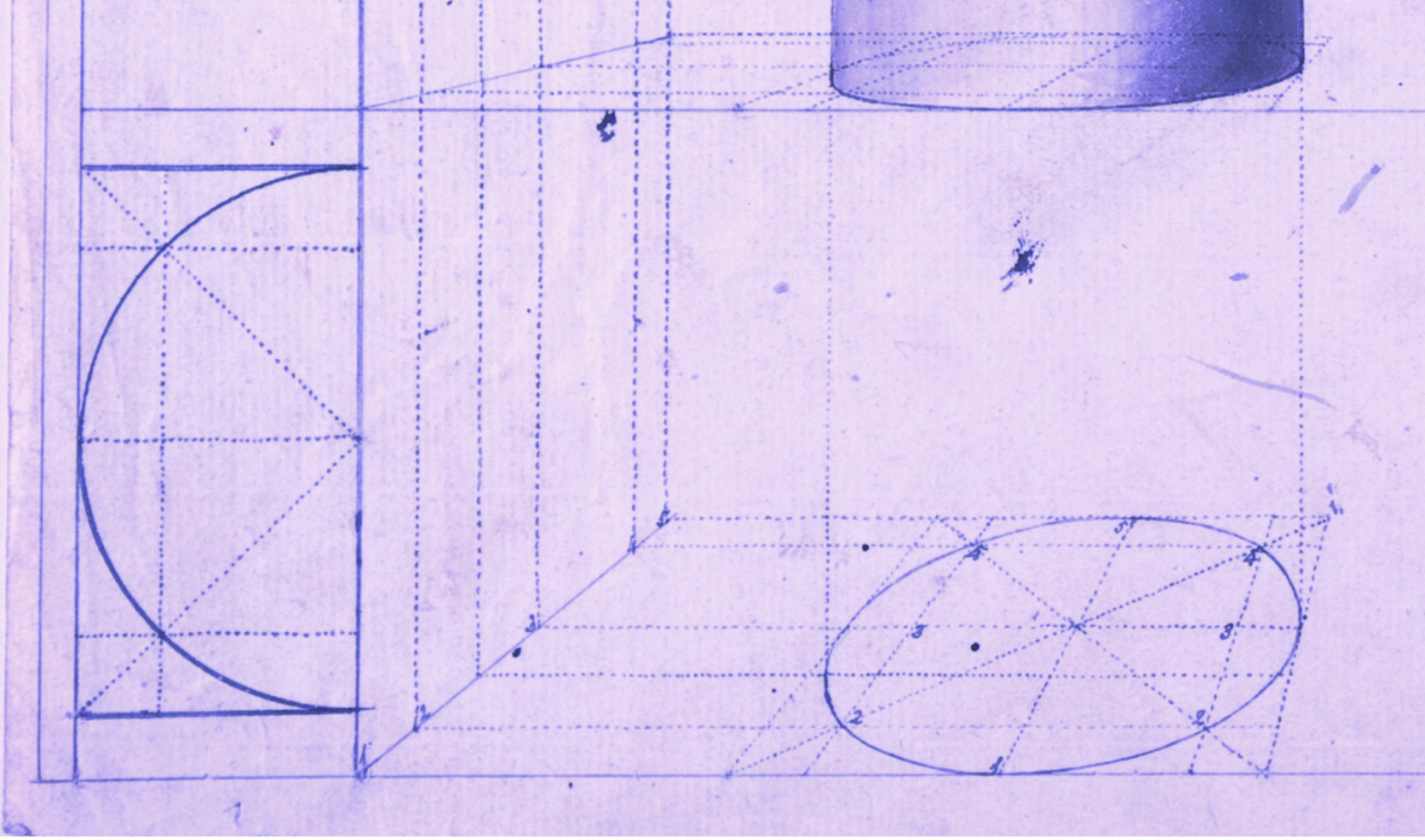
---

[6] Despite revenue growth, profit often remains elusive; *PitchBook Profile - OpenAI*, https://content.pitchbook.com/share/quick-link/profile/9374fa6e-c6a6-4a7f-8402-fbc780b08e48?hash=8fac7125145ad4bf2fc8a3a6fc1239b49457cc22a3ed196a5020b3b081e26fa9 (last visited June 23, 2025) (showing that OpenAI lost $5 billion in 2025, despite drawing a revenue upwards of $3.7 billion and a net profit margin of -134% in 2024); Cory Weinberg, *OpenAI Projections Imply Losses Tripling to $14 Billion in 2026*, Information (Oct. 9, 2024), https://www.theinformation.com/articles/openai-projections-imply-losses-tripling-to-14-billion-in-2026 (last visited Oct. 27, 2025); *Will OpenAI Ever Make Real Money?*, Economist (May 15, 2025), https://www.economist.com/business/2025/05/15/will-openai-ever-make-real-money (last visited June 2, 2025); Mike Isaac & Erin Griffith, *OpenAI Is Growing Fast and Burning Through Piles of Money*, N.Y. Times (Sep. 27, 2024), https://www.nytimes.com/2024/09/27/technology/openai-chatgpt-investors-funding.html; Berber Jin, *Mind-Blowing Growth Is About to Propel Anthropic Into Its First Profitable Quarter*, Wall St. J. (May 20, 2026), https://www.wsj.com/tech/ai/mind-blowing-growth-is-about-to-propel-anthropic-into-its-first-profitable-quarter-7edbf2f4 (describing unaudited accounting projects in Anthropic's first profitable quarter, but noting that profits are not expected to last given the company's infrastructure spending); Anthropic, PBC, Amendment No. 1 to Registration Statement (Form S-1/A) 23 (Nov. 2, 2026), https://www.sec.gov/Archives/edgar/data/2130357/000162828026068676/anthropicpbc-sx1ax1.htm (disclosing a $1.26 billion operating loss on $33.2 billion revenue in the first nine months of 2026).

[7] *Artificial Intelligence - Worldwide | Market Forecast*, Statista, https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide/ (last visited July 14, 2025) [https://web.archive.org/web/20250715163558/https://www.statista.com/outlook/tmo/artificial-intelligence/worldwide] (estimating global growth up to $1 trillion by 2031, including consumer facing and enterprise AI products).

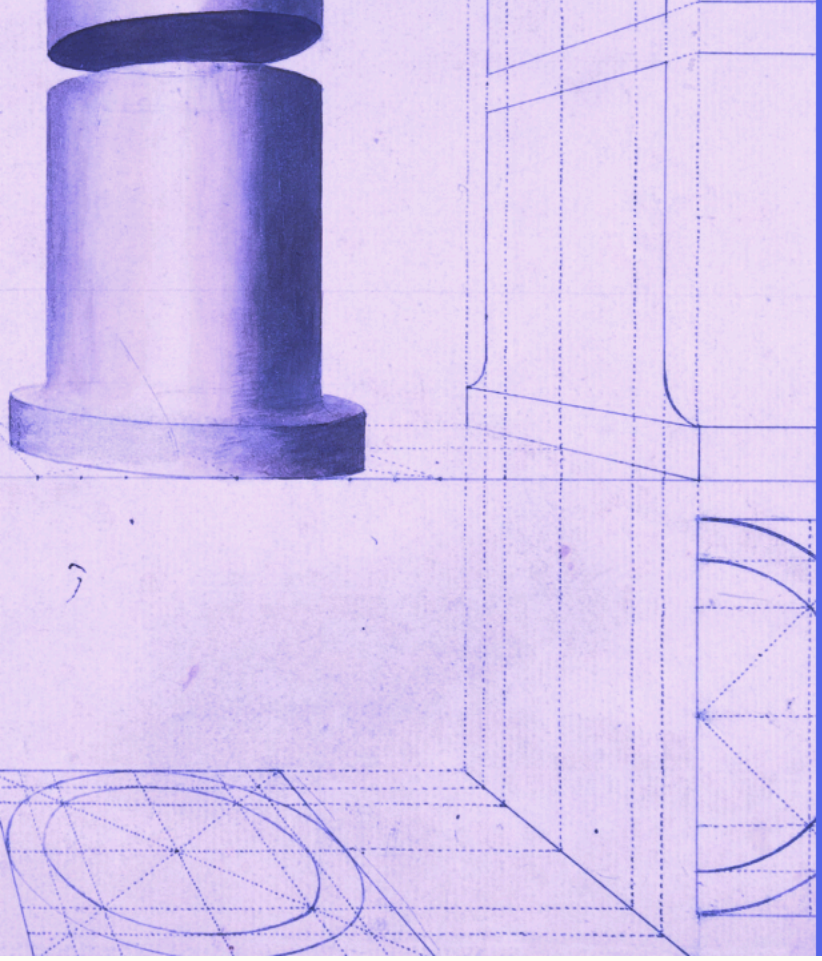
This report assesses how various business structures fare in practice in terms of their stated commitments to benefit humanity. Though the contours of the mission of AI companies remain vague, morph over time and vary across companies, at a high level they all signal a commitment to public interest, as opposed to the traditional goal of corporations which consists of maximizing profit for shareholders. Taking that promise as more than newspeak, the case studies below assess whether variations among corporate structures have any bearing on companies' ability to fulfill their missions. The aim is to surface correlations between how companies are structured and how they behave, comparing how different corporate structures enable them to carry out their stated mission. Confounding variables such as founder culture, investor preferences, and market position stand in the way of a causal relationship between corporate structure and behavior. Rather, the report starts a conversation about corporate structure as one factor shaping how AI labs walk the talk on their mission, and highlights corporate law levers to hold AI labs to their word.

Part I explains the legal business structures of Google DeepMind, OpenAI and Anthropic, outlining their evolution over time. Part II assesses how each company behaves in practice, based on their policy positions on AI regulation, internal safety practices and military contracts. Given the pace of events, the report tracks corporate conduct through February 2026. In fact, the shapeshifting of corporate structures (e.g., OpenAI pivoting from a non-profit to a 'capped-profit' and then to a public benefit corporation) is itself an important insight, suggesting a strategic instability that impedes analysis. Nevertheless, these case studies give a sense of how AI companies negotiate tensions between their public mission and self-interested considerations. Part III proposes avenues to demand companies follow through on their mission.



Part I. Taking Stock of Business Structures

AI companies are adept at crafting bespoke legal structures to hardcode public interest in their DNA. Yet as AI went from proof of concept to viable product, they have all diluted such commitments. Over time, AI labs have tended to adapt their corporate form to be more responsive to investors, citing the need to secure massive investments to train and run ever larger frontier models. Google absorbed DeepMind, and, along the way, shed plans for external oversight. Originally a non-profit, OpenAI's structural changes introduced powerful financial considerations. Anthropic's structure promotes public interest governance, but it remains brittle as profit-driven shareholders retain the ability to roll back guardrails. After introducing basic concepts of business organization, this section details the structure of the leading AI companies under review.



Overview

There are two important distinctions when it comes to AI companies. The first pertains to their status as for-profit or non-profit entities. The second is the specific legal vehicle under which they operate.

1. For-Profit /Non-Profit Status

For-profit and non-profit entities operate under different legal frameworks. **For-profit** entities prioritize generating revenue for their owners. **Non-profit** entities pursue broader public interest, in return for which they receive significant tax advantages.[8] They are often called 501(c)(3)s, a shorthand for the relevant federal tax code provision exemption. Non-profits must act consistently with their charitable purpose and use their assets in pursuance of their mission. The last section lays out in greater detail the protection of the charitable purpose and assets of non-profit organizations.

2. Legal Vehicle

AI companies have to choose a specific legal vehicle to carry out their activities. Corporations, limited liability companies (LLCs) and limited partnerships (LPs) involve different goals and rules in terms of control, ownership, transparency and liability.[9]

---

[8] The variations of fiscal treatment among these various legal vehicles are beyond the scope of this report.

[9] Since Delaware is the jurisdiction of choice for incorporation for most U.S. businesses, including AI companies, the descriptions below reflect Delaware rules unless noted otherwise. *See* Del. Code Ann. tit. 6, § 15-19, tit. 8 (2026). Other than the mission as stated in tax exemptions applications, the tax implications of these various legal vehicles are beyond the scope of this report.

Corporations are the paradigmatic vehicle to organize commercial activities. The primary goal of a corporation is to generate profit for its shareholders.[10] In terms of control, a board of directors elected by shareholders oversees management. These shareholders have a stake in the corporation, which sells them stocks to raise capital. The board of directors approves distribution of profits to shareholders, in the form of dividends. Corporations publicly trading stock have to disclose their governance structure as well as their financial performance.[11] These disclosures are meant to protect investors by giving them enough information to make informed decisions. In terms of liability, corporations are legally distinct entities separate from their shareholders. Known as the "corporate veil," this legal separation isolates shareholders' assets from corporate obligations. The personal assets of shareholders are protected and their losses are capped at their investment.

- ➔ Alphabet, Google's parent company, is a corporation.
- ➔ OpenAI's original non-profit entity was a particular form of "non-stock" corporation, collecting donations instead of raising capital by selling shares.

[10] Milton Friedman, *A Friedman Doctrine— The Social Responsibility of Business Is to Increase Its Profits*, N.Y. Times (Sep. 13, 1970), https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html (stating that "there is one and only one social responsibility of business—to use its resources and engage in activities designed to increase its profits"); Stephen M. Bainbridge, The Profit Motive: Defending Shareholder Value Maximization (2023). Even in "softer" stakeholderism, considering other interests is still ultimately about profitability. *See, e.g.*, *Larry Fink's Annual 2022 Letter to CEOs*, BlackRock, https://www.blackrock.com/corporate/investor-relations/2022-larry-fink-ceo-letter (last visited July 14, 2025) ("Make no mistake, the fair pursuit of profit is still what animates markets; and long-term profitability is the measure by which markets will ultimately determine your company's success.") (cited in *Developments in the Law—Amoral Drift in AI Corporate Governance*, Harv. L. Rev. 1633, 1639 n.32 (2025), https://harvardlawreview.org/print/vol-138/amoral-drift-in-ai-corporate-governance/). For a list of controlling cases in Delaware, the leading US corporate law jurisdiction, *see* Ann M. Lipton, *Will the Real Shareholder Primacy Please Stand Up?*, Harv. L. Rev. 1584, 1592–93 n.55 (2024), https://harvardlawreview.org/print/vol-137/will-the-real-shareholder-primacy-please-stand-up/. *See also infra* note 31 on the limits of shareholder primacy and the broader welfare of shareholders beyond profits.

[11] *Exchange Act Reporting and Registration*, U.S. Securities and Exchange Commission, https://www.sec.gov/resources-small-businesses/going-public/exchange-act-reporting-registration.

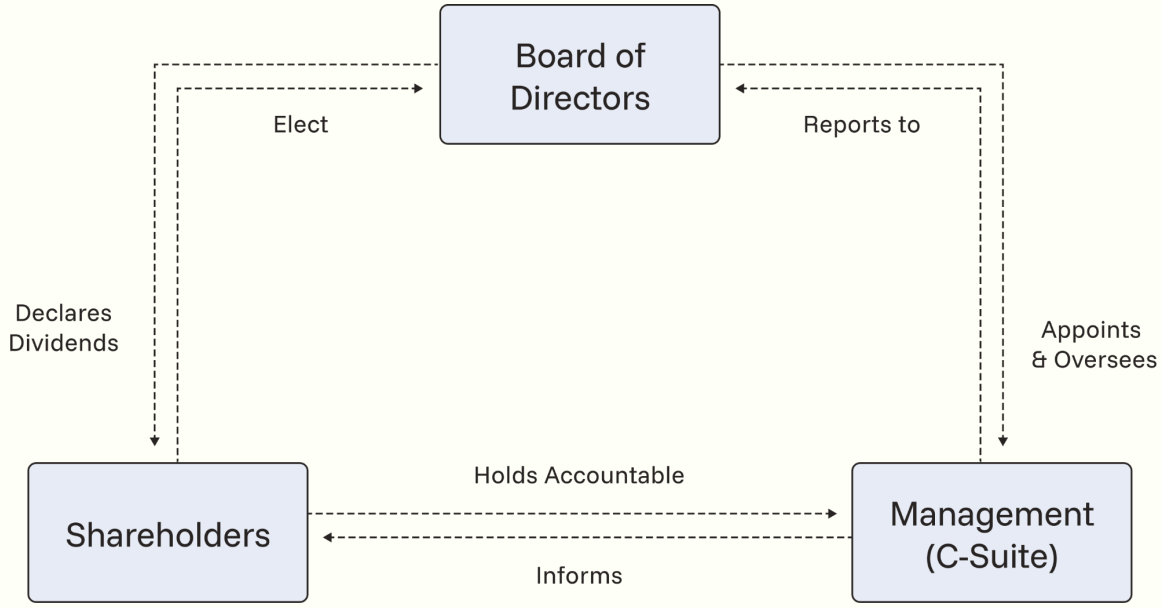


Figure 1: Traditional Corporation

Public Benefit Corporations (PBCs) seek to combine the profit-making goals of conventional corporations with social goals. Over 40 U.S. states have enacted statutory frameworks to allow businesses to structure as PBCs. PBCs share the same control, ownership and limited liability as traditional corporations, but require directors to balance profit with purpose. In practice, PBC directors can, but don't have to, prioritize the mission. Because directors have significant leeway in how they approach their mandate, PBCs are best understood as enabling rather than requiring a company to pursue social goals. In addition to the disclosure requirements of a traditional corporation, PBCs have to report how they achieve their stated public interest objectives. Delaware, currently the primary jurisdiction for corporation law in the U.S., allows PBCs to keep this report confidential.

➔ OpenAI and Anthropic currently operate as PBCs.

Limited Liability Companies (LLCs) combine flexible management options with liability protection for their owners, called members. Managers run the AI LLCs under review, and a confidential operating agreement outlines the governance structure and the relationships between managers and members. As with corporations, LLCs are separate legal entities, such that neither managers nor members are responsible for the company's obligations beyond their investment.

➔ Google is an LLC, owned and controlled by the Alphabet Corporation.

➔ OpenAI set up several LLCs during its 2019-2025 "capped-profit" phase. These LLCs primarily served two functions: they allowed the non-profit arm to maintain formal control over the

for-profit arms, and they raised capital from major investors like Microsoft.

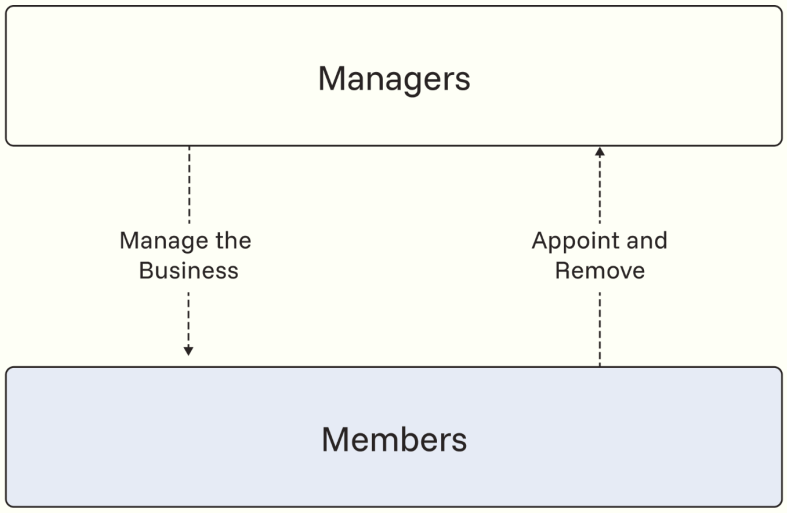


Figure 2: Limited Liability Corporation

Limited Partnerships (LPs) enable a group of people to own a business together and share profits. There are two types of partners with distinct roles and levels of exposure to liability. General partners actively manage the business, whereas limited partners are passive investors. A confidential limited partnership agreement outlines the rights, responsibilities, and profit-sharing structure of the partnership. Unlike corporations and LLCs, LPs are not distinct legal entities. General partners bear full personal responsibility for the LP they manage, whereas limited partners typically only risk their investment. In other words, creditors can go after a general partner's assets, but they cannot do so with limited partners.

➔ OpenAI created a limited partnership to raise capital and offer equity to employees. Investors and employees were limited partners, with the non-profit entity acting as the general partner.

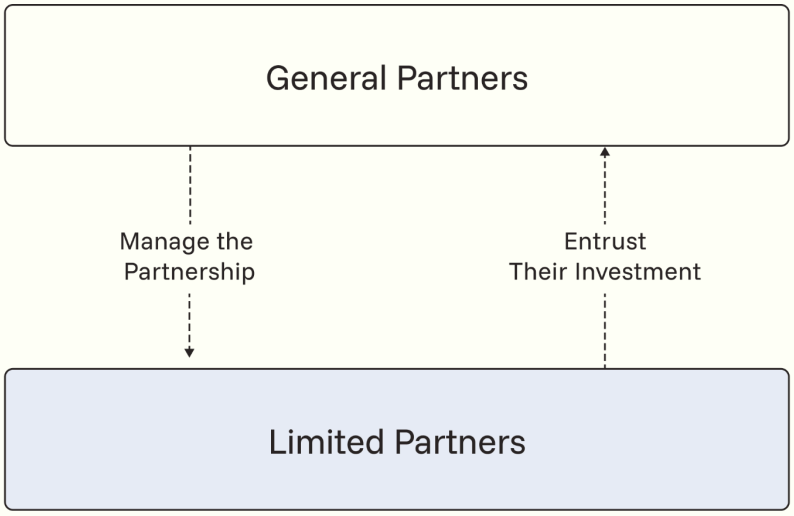
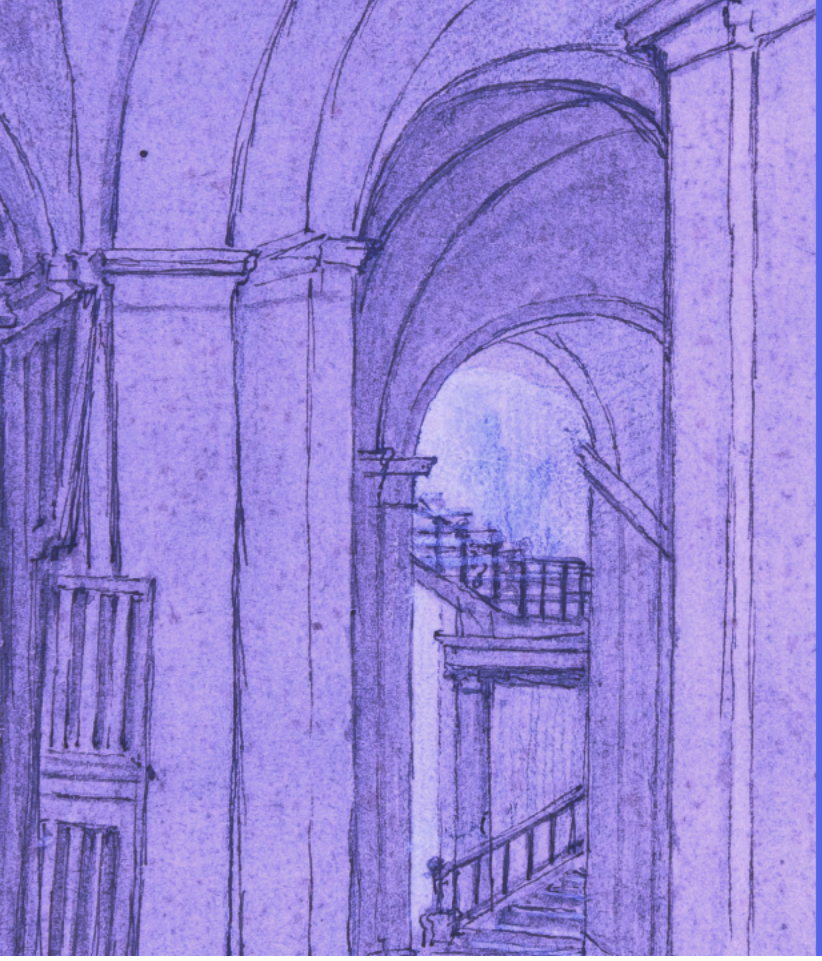


Figure 3: Limited Partnership

A for-profit corporation like Alphabet (Google's parent company) primarily exists to maximize profit, and shareholders have various avenues to ensure directors steer the company accordingly.[12] It provides a baseline to assess whether the innovative business structures of OpenAI and Anthropic yield a better result in terms of aligning mission and action. Departing from the traditional business model, these frontier labs have crafted bespoke legal structures with a dual purpose: attract investment while pursuing their mission to benefit humanity. All three companies have evolved their structure over time, as summarized below.

[12] These include voting out directors and actions to enforce their duties to act in the corporation and the shareholder's interests. *See* Practical Law Corporate & Securities, *Fiduciary Duties of the Board of Directors* (Westlaw) (2023), https://law.stanford.edu/wp-content/uploads/2023/01/Fiduciary-Duties-of-the-Board-of-Directors.pdf. For more on fiduciary duties, see info box below "What are fiduciary duties?".



DeepMind/Google

DeepMind began as an independent startup striving to "solve intelligence" and "use it to solve everything else."[13] It now leads Google's AI activities including its flagship chatbot and model Gemini. The relationship between the research lab and its acquirer Google has been one of push and pull. Various mutations in DeepMind's corporate structure evidenced a desire to maintain independence, but the gravitational pull of the product-driven parent company ultimately prevailed. Over the years, DeepMind's independence has gradually receded as it was amalgamated into the Google/Alphabet ecosystem.

1. From Private Startup to Google Acquisition

A research pioneer founded in the UK in 2010, DeepMind began as a UK private limited company.[14] Similar to a Delaware corporation, this vehicle offers limited liability protecting shareholders, who in turn elect a board of directors in charge of governance. After declining an offer from Facebook and Elon Musk, it agreed to a Google acquisition in 2014 in return for an ethics board preserving the

[13] Hal Hodson, *DeepMind and Google: The Battle to Control Artificial Intelligence*, Economist (Mar. 1, 2019), https://www.economist.com/1843/2019/03/01/deepmind-and-google-the-battle-to-control-artificial-intelligence; *see also* Parmy Olson, Supremacy: AI, ChatGPT, and the Race That Will Change the World 52-53 (2024).

[14] *DeepMind Certificate of Incorporation* (2010), https://find-and-update.company-information.service.gov.uk/company/07386350/filing-history. Early investors included Elon Musk, Peter Thiel, Jaan Tallinn (Skype), Yoshua Bengio and Geoffrey Hinton. For a full list, see *DeepMind Technologies Limited - Annual Return, Company Number 07386350* (2014), https://find-and-update.company-information.service.gov.uk/company/07386350/filing-history; Olson, *supra* note 13 57 (reporting Peter Thiel's initial £1.4m investment).

company's mission and control over artificial general intelligence (AGI).[15] Google dissolved the ethics board after a single meeting.[16]

DeepMind has struggled for more independence ever since. When Google floated a preliminary offer for a partial spin-out,[17] it explored a bespoke corporate structure dubbed "Global Interest Company" with two poles of governance: an operating board overseeing everyday activities and a separate board of trustees ensuring compliance with social and ethical requirements.[18] In the spirit of the UN, this Global Interest Company prototype would have acted in the general interest and licensed its business-relevant technology to Google, with a carve-out against military applications.[19] The Global Interest Company never materialized, and neither did a more modest option to become a company limited by guarantee, a legal entity commonly used by non-profit organizations in the UK.[20]

Instead, Google pulled in DeepMind closer to product development. As the Google ecosystem went through a reorganization in 2015 putting Alphabet at the apex of its conglomerate, DeepMind eventually operated as a subset of Google. In 2023, Alphabet merged Google Brain (the company's AI research team) and DeepMind into a single unit operating under the Google umbrella,[21] chipping away at DeepMind's independence.[22] This reconfiguration shifted the research lab from a marginal bet in Alphabet's portfolio, along with self-driving car company Waymo and other moonshots not expected to

---

[15] *See* Steven Levy, *The Deep Mind of Demis Hassabis*, Wired, Jan. 16, 2015, https://www.wired.com/2015/01/the-deep-mind-of-demis-hassabis/ (reporting the statement of DeepMind's founder, Demis Hassabis, that the ethics board was part of the acquisition deal); *see also* Olson, *supra* note 13, at 63–64, 77, 119 (reporting that DeepMind left some $150 million on the table when it turned down Facebook's acquisition offer in order to keep the ethics board on offer in Google's proposed acquisition); Samuel Gibbs, *Google Buys UK Artificial Intelligence Startup Deepmind for £400m*, Guardian (Jan. 27, 2014), https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind (noting that the sale value was roughly equivalent to $650 million at that time); Amir Efrati, *Google Beat Facebook for DeepMind, Creates Ethics Board*, Information (Jan. 26, 2014), https://www.theinformation.com/articles/google-beat-facebook-for-deepmind-creates-ethics-board. For simplicity, we refer to the Google DeepMind division of Google within Alphabet as simply Google unless the context warrants specification.

[16] Cade Metz et al., *Ego, Fear and Money: How the A.I. Fuse Was Lit*, N.Y. Times (Dec. 3, 2023), https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html; Olson, *supra* note 13, at 81.

[17] Olson, *supra* note 13, at 82, 108 (describing how the offer would endow DeepMind with a $15 billion fund over ten years and a separate board of trustees while allowing it to retain some Google/Alphabet ownership).

[18] *Id.* at 109; *see also* Martin Coulter & Hugh Langley, *DeepMind Spent Years Trying to Break Away from Google. Insiders Detail a Secret Plot Sparked by Distrust and Driven by Fears the Search Giant Would Sell Its AI to the Military*, Bus. Insider (Sep. 11, 2021), https://www.businessinsider.com/deepmind-secret-plot-break-away-from-google-project-watermelon-mario-2021-9 (last visited July 11, 2025).

[19] *See* Olson, *supra* note 13, at 109, 119. This information could not be independently verified.

[20] Coulter & Langley, *supra* note 18; Olson, *supra* note 13, at 185; Brenda Hannigan, Company Law 17–19 (1st ed. 2003).

[21] *Google DeepMind: Bringing Together Two World-Class AI Teams*, Google (Apr. 20, 2023), https://blog.google/technology/ai/april-ai-update/.

[22] DeepMind Technologies Limited - Annual Return, company number 07386350, *supra* note 14, at 8 (listing Google as the sole shareholder in 2014); Mike Nudelman & Rob Price, *One Chart That Explains Alphabet, Google's Parent Company*, Bus. Insider (July 23, 2016), https://www.businessinsider.com/chart-alphabet-google-parent-company-deepmind-gv-x-nest-etc-2016-7 (last visited July 10, 2025) (describing how DeepMind previously sat under Google and was then spun out as a separate entity under the Alphabet umbrella); *Audited Financial Statements DeepMind Technologies Limited* (2015), https://find-and-update.company-information.service.gov.uk/company/07386350/filing-history/MzE1NzU0Nzk4OGFkaXF6a2N4/document?format=pdf ("The Company is wholly owned by Google Ireland Holdings, which is incorporated in Ireland. Alphabet Inc. is the ultimate parent undertaking and is a Company incorporated in the United States of America."); *DeepMind Technologies Limited - Confirmation Statement - Company Number 07386350* (2016) (listing Alphabet as a 'person with significant control', i.e., 75% or more shares).

generate revenue, to the core product-driven business of Alphabet.[23] Google further consolidated its research and DeepMind division in 2024,[24] and moved teams working on Gemini under the single Google DeepMind umbrella in 2025.[25] Though DeepMind gained an ethics council in the process, Google personnel exclusively staffed it.[26] DeepMind employees had raised concerns that "their technology would be too powerful to be held by a private company, so it needs to be housed in some other legal entity detached from shareholder interest," but a corresponding corporate structure never came to pass.[27]

[23] Katie Collins, *Dip into DeepMind's 3D Playground Where AI Comes to Learn*, CNET (Dec. 5, 2016), https://www.cnet.com/culture/google-alphabet-deepmind-lab-3d-navigation-for-ai-software/ (framing DeepMind as a moonshot project with no expectation of short term profitability); Jeremy Kahn, *The Google Brain-DeepMind Merger Is Probably Good for Google. It Might Not Be for Us*, Fortune (Apr. 28, 2023), https://fortune.com/europe/2023/04/28/the-google-brain-deepmind-merger-alphabet-pichai-risks-eye-on-a-i/ (reporting the shift in Alphabet's SEC filings moving DeepMind from the "other bet" category to the main profit and losses statement, reflecting its growing importance to the business).

[24] *Google Consolidates Its DeepMind and Research Teams amid AI Push*, Reuters (Apr. 18, 2024), https://www.reuters.com/technology/google-consolidates-its-deepmind-research-teams-amid-ai-push-2024-04-18/.

[25] Kyle Wiggers, *Google Folds More AI Teams into DeepMind to "Accelerate the Research to Developer Pipeline,"* TechCrunch (Jan. 9, 2025), https://techcrunch.com/2025/01/09/google-folds-more-ai-teams-into-deepmind-to-accelerate-the-research-to-developer-pipeline/.

[26] Laura Weidinger et al., *Holistic Safety and Responsibility Evaluations of Advanced AI Models,* Apr. 22, 2024, at 6, http://arxiv.org/abs/2404.14068 (describing the Responsibility and Safety Council as made up of senior Google DeepMind experts); *Responsibility & Safety*, Google DeepMind (Nov. 2025), https://deepmind.google/responsibility-and-safety/ (describing Responsibility and Safety Council as an internal review group); Olson, *supra* note 13, at 189; DeepMind Safety Research, *AGI Safety and Alignment at Google DeepMind: A Summary of Recent Work*, Medium (Oct. 18, 2024), https://deepmindsafetyresearch.medium.com/agi-safety-and-alignment-at-google-deepmind-a-summary-of-recent-work-8e600aca582a (describing the internal team in the AGI Safety Council).

[27] Coulter & Langley, *supra* note 18.

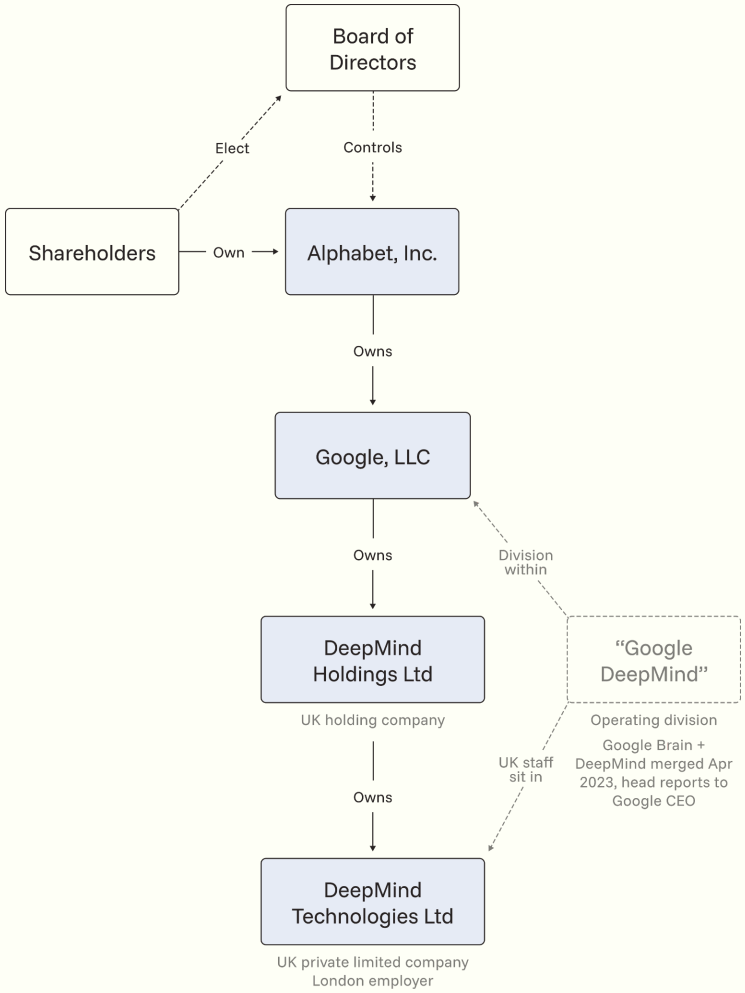


Figure 4: Google DeepMind is part of Google LLC under the Alphabet umbrella.

In August 2026, Google's chief scientist and longtime employee Jeff Dean stepped down, with DeepMind founder Demis Hassabis moved from Google DeepMind CEO to Chair of the unit and to the newly created role of Chief Scientist of Alphabet. A Senior Vice President oversees Google DeepMind's day-to-day activities and reports directly to Google's CEO. Some employees have reportedly noted that the reshuffle further shifted Google DeepMind's center of gravity from its original London offices to Alphabet's headquarters in Mountain View, CA.[28]

[28] Beatrice Nolan, *How Stalled Models, Missed Deadlines, and Staff Burnout Led to the Unraveling of Google's DeepMind*, Fortune (Aug. 10, 2026), https://fortune.com/2026/08/10/how-stalled-models-missed-deadlines-and-staff-burnout-lead-to-the-unraveling-of-googles-deepmind/.

2. A Cog in the Alphabet Machine

With DeepMind now fully assimilated into Google, the parent company Alphabet has full control and ownership. As seen in figures 2 and 3, shareholders elect the board of Alphabet, a publicly listed corporation with a dual-class share structure giving founders the majority of voting power to elect the board, on which they also sit.[29] Directors are required to own a large portion of stock to align their interests with those of shareholders.[30]

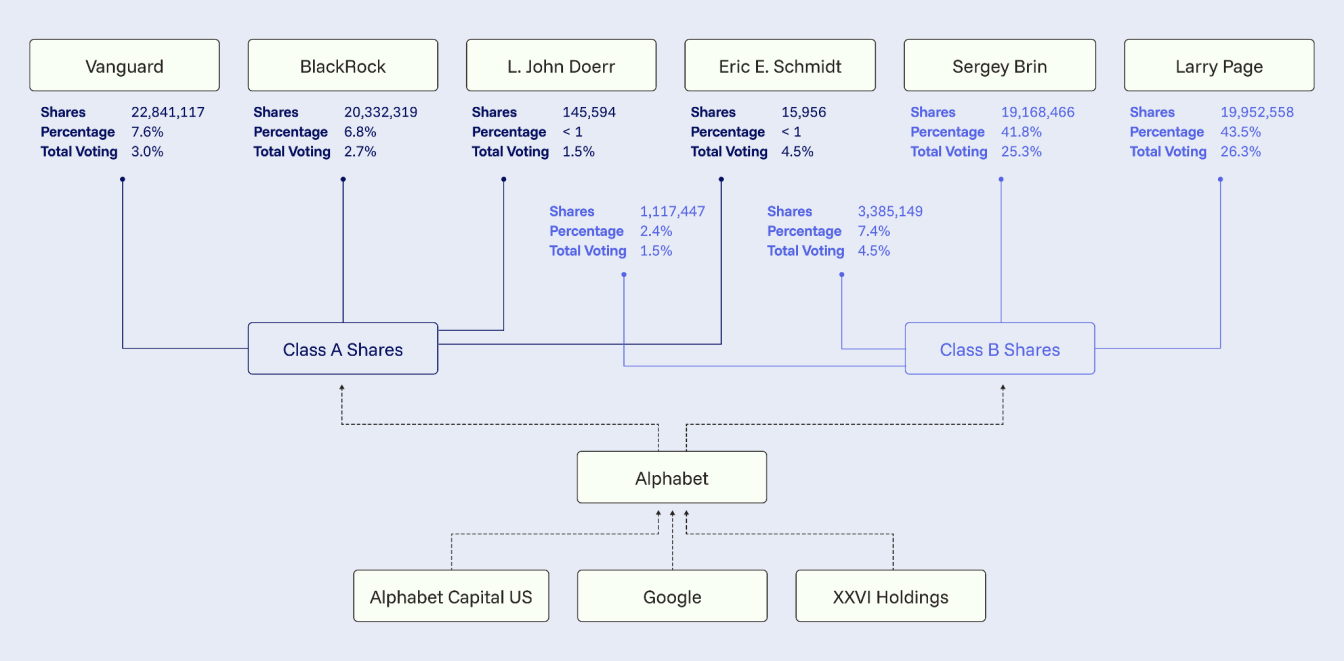


Figure 5: Under Alphabet's dual-class share structure, founders retain control despite owning a smaller percentage of the company. This Figure shows major institutional shareholders in 2022. Class A shares confer one vote per share, class B shares 10 votes per share. Source: Lexchart.com

Alphabet operates under the typical shareholder-centric paradigm of for-profit corporations. Though academics debate whether and how traditional corporations can pursue objectives beyond profit, the

[29] Alphabet, *Amended Restated Certificate of Incorporation*, Alphabet Investor Relations, Art. IV (July 5, 2022), http://abc.xyz/investor/board-and-governance/certificate-of-incorporation (giving 1 vote per share for Class A shares for regular investors, 10 votes per shares for Class B shares for founders, and no vote per Class C share for employees); Art. IX on election and removal process; *see also Alphabet Bylaws*, Alphabet Investor Relations, Art. 313 (July 5, 2022), http://abc.xyz/investor/board-and-governance/bylaws; *Corporate Governance Guidelines*, Alphabet Investor Relations (July 5, 2022), http://abc.xyz/investor/board-and-governance/corporate-governance-guidelines (outlining the election process). For an overview of voting power per investor, *see Alphabet (Google) Company Structure Chart with Subsidiaries*, Lexchart, Feb. 15, 2022, https://lexchart.com/org-charts/alphabet-google/.

[30] Corporate Governance Guidelines, *supra* note 29 Art. IV 9 (directors have to own shares worth a least a million dollars to "more closely align the interests of our directors and senior management with those of our stockholders").

prevailing view enforced by courts remains maximizing shareholder value.[31] What's for sure is that unlike OpenAI or Anthropic, Google lacks a clear public-benefit mandate. Instead, it professes public responsibility through self-imposed policies. In 2018, Google responded to internal criticism by publishing a set of principles committing to abstain from certain AI applications such as weapons and surveillance.[32] However, it reversed course on these non-binding pledges in 2025.[33] In terms of voluntary oversight, its 2019 external AI ethics advisory council was dissolved within a week amid controversy over the transphobic, xenophobic views of one of its members.[34] In practice, Alphabet relies on managerial discretion to honor ethical commitments,[35] and traditional fiduciary duties to maximize shareholder value legally govern its decision-making. Ultimately, Alphabet treats controversial AI activities as reputational risk.[36] Under federal securities laws, companies publicly trading stock have a duty to disclose material risks. Its latest filing with the Securities and Exchange Commission (the U.S. federal agency overseeing publicly listed companies) notes that "[i]ssues in the development and use of AI may result in reputational harm and increased liability exposure."[37] The last section explores shareholder protection as a lever for accountability in greater detail.

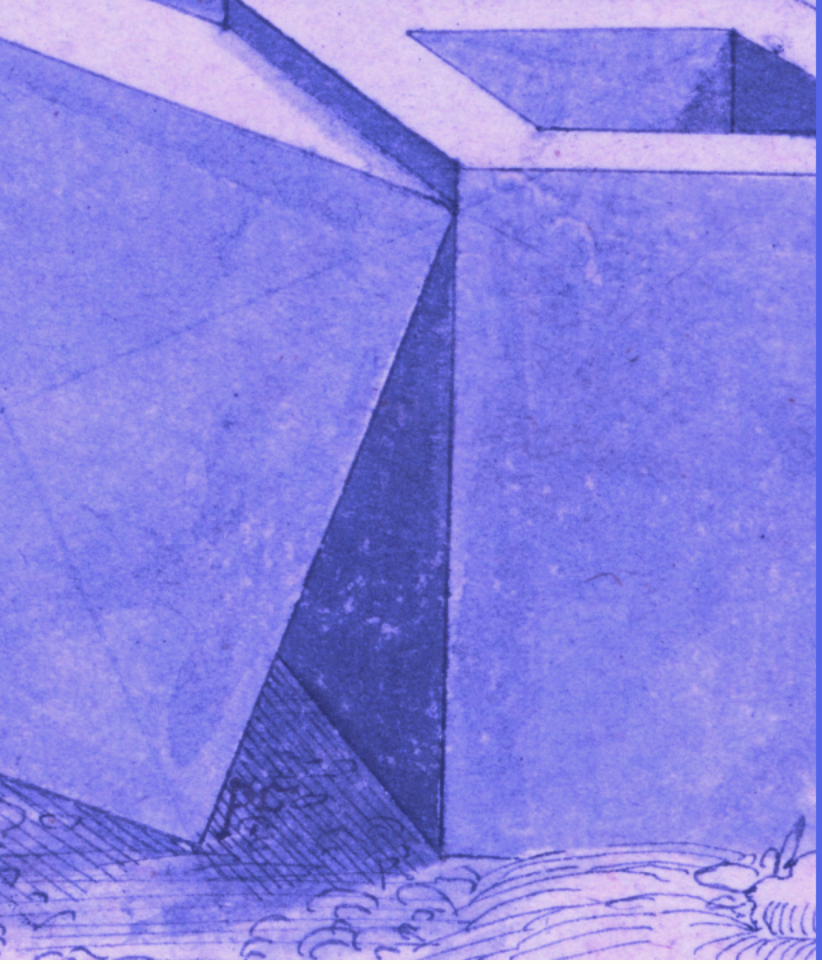
What are fiduciary duties?

Fiduciary duty is a legal term of art requiring a party (the fiduciary) to act in the best interests of another (the beneficiary). In the corporate law context, the people managing businesses owe fiduciary duties to investors, be they partners of a limited partnership, members of limited liability corporations, or shareholders of a corporation. There are three main duties: loyalty,

[31] Emilie Aguirre, *Beyond Profit*, 54 U.C. Davis L. Rev. 2077, 2100–01 (2021); *Dodge v. Ford Motor Co.*, 170 N.W. 668, 684 (Mich. 1919) ("A business corporation is organized and carried on primarily for the profit of the stockholders."); Robert Rhee, *A Legal Theory of Shareholder Primacy*, 122 Minn. L. Rev. 1951 (2018), https://scholarship.law.umn.edu/mlr/122; *but see* D. Gordon Smith, *The Shareholder Primacy Norm*, 23 J. CORP. L. 277 (1998) (enforcing shareholder primacy on publicly listed companies is difficult in practice); Oliver Hart & Luigi Zingales, Companies Should Maximize Shareholder Welfare Not Market Value (Aug. 1, 2017) (unpublished manuscript), https://papers.ssrn.com/abstract=3004794 (arguing that shareholder welfare should go beyond profits) (cited in Lipton, *supra* note 10, at 1638).
[32] Pichai, *supra* note 4.
[33] *AI Principles*, Google AI, https://ai.google/principles/ [https://perma.cc/Q88G-LU2N] (last visited May 22, 2025); Paresh Dave, *Google Lifts a Ban on Using Its AI for Weapons and Surveillance*, Wired (Feb, 4, 2025), https://www.wired.com/story/google-responsible-ai-principles/ (last visited July 17, 2025).
[34] Nick Statt, *Google Dissolves AI Ethics Board Just One Week after Forming It*, Verge (Apr. 4, 2019), https://www.theverge.com/2019/4/4/18296113/google-ai-ethics-board-ends-controversy-kay-coles-james-heritage-foundation.
[35] Nitasha Tiku, *A Top Google Exec Pushed the Company to Commit to Human Rights. Then Google Pushed Him out, He Says*, Wash. Post (Jan. 2, 2020), https://www.washingtonpost.com/technology/2020/01/02/top-google-exec-pushed-company-commit-human-rights-then-google-pushed-him-out-he-says/ (recounting dwindling human rights orientation in executive decisions).
[36] Sam Biddle, *Google Worried It Couldn't Control How Israel Uses Project Nimbus, Files Reveal*, Intercept (May 12, 2025), https://theintercept.com/2025/05/12/google-nimbus-israel-military-ai-human-rights/ (reporting that a $3.3 billion government cloud contract raised concerns about reputational harm from complicity with human rights violations).
[37] Alphabet Inc., *Form 10-K*, 16 (2025), https://www.sec.gov/Archives/edgar/data/1652044/000165204425000014/goog-20241231.htm.

good faith and care.[38] Loyalty requires fiduciaries to act in the best interest of the company. Good faith involves acting honestly, without intentional misconduct. The duty of care focuses on informed decision-making, but courts are fairly deferential. Under the business judgment rule, they will only punish gross negligence of the duty of care.[39]

[38] Julian Velasco, *Fiduciary Principles in Corporate Law*, *in* The Oxford Handbook of Fiduciary Law 0, 71 (Evan J. Criddle, Paul B. Miller, & Robert H. Sitkoff eds., 2019), https://doi.org/10.1093/oxfordhb/9780190634100.013.4 (noting that the duty of good faith is subsumed into the duty of loyalty in Delaware).
[39] *Id.* at 69.



OpenAI

OpenAI is the company behind some of the most prominent AI products such as ChatGPT and former video generating application Sora. Initially set up as a single non-profit, it then nested a fleet of for-profit entities under an overarching non-profit corporation, only to morph once again into a single public benefit corporation nominally overseen by a non-profit foundation. The company started off with an explicit mission to benefit the public, "not organized for the private gain of any person."[40] Its tax-exemption application frames it as a "nonprofit artificial intelligence ("AI") scientific research organization" aiming to "engage in research activities that advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return."[41] Its non-binding 2018 Charter further commits to a "fiduciary duty to humanity," noting concerns about a competitive race to the bottom and power concentration issues in the field.[42] However, the company's corporate structure has significantly evolved over time, calling into question its ability to embody these ideals. Citing the need to raise more capital as its technology gained traction from proof-of-concept to

[40] OpenAI, *Certificate of Incorporation and Bylaws* 1 (2015), https://www.documentcloud.org/documents/25197523-openai-application-for-tax-exempt-status/.
[41] OpenAI, *Internal Revenue Service Form 1023, supra* note 4, at 30.
[42] OpenAI Charter, *supra* note 2 (defining AGI as "highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity"). For a critique of AGI as an "aspirational technological construct" and a fuller genealogy of the company, see Brian Merchant, AI Generated Business: The Rise of AGI and the Rush to Find a Working Revenue Model 3 (2024), https://ainowinstitute.org/publications/ai-generated-business.

overnight success with ChatGPT,[43] it altered its business model and corresponding legal structure multiple times. Three main phases stand out: a non-profit startup lab to perform open research, a joint venture hybrid to raise private capital, and a public benefit corporation to combine profit and purpose.[44] The final section of this report documents how attempts to move away from the original form were met with resistance.

1. The Genesis of OpenAI: Non-profit Research Lab

OpenAI started as a non-profit, mission-driven countervailing force to Google/DeepMind in 2015.[45] Worried about the threat of superhuman machine intelligence, especially if led by profit-driven, monopolistic Google, its founders sought an alternative way to build safe and beneficial AI technology.[46] Its founders include Sam Altman, then president of tech startup accelerator Y Combinator; Greg Brockman, formerly chief technology officer of online payment processing platform Stripe; Elon Musk, CEO of Tesla and SpaceX; and Ilya Sutskever, renowned research scientist recruited from Google.

By organizing as a non-stock corporation with a non-profit status, OpenAI initially operated on donations, tax-free and shielded from the pressure to turn profits.[47] OpenAI, Inc. didn't issue shares to raise capital, so it didn't have to seek profit in order to pay dividends to shareholders. Instead, this "non-stock" corporation collected donations without an expectation of return on investment. CEO Sam Altman stated that "[w]e don't ever want to be making decisions to benefit shareholders. The only people we want to be accountable to is humanity as a whole."[48] The initial funding pledges from its

[43] Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base*, Reuters (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (last visited July 7, 2025) (reporting that ChatGPT had reached 100 million users in just two months). Threads has since surpassed ChatGPT, but it drew from Instagram's user base, such that the comparison is of limited use. *See* Siladitya Ray, *Threads Now Fastest-Growing App In History—With 100 Million Users In Just Five Days*, Forbes (July 10, 2023), https://www.forbes.com/sites/siladityaray/2023/07/10/with-100-million-users-in-five-days-threads-is-the-fastest-growing-app-in-history/ (last visited July 7, 2025).

[44] On the transition to a capped-profit joint venture, see *OpenAI LP*, OpenAI, (2019), https://openai.com/index/openai-lp/ ("We want to increase our ability to raise capital while still serving our mission, and no pre-existing legal structure we know of strikes the right balance."). On the move to a public benefit corporation, see *Evolving OpenAI's Structure*, OpenAI, (May 5, 2025), https://openai.com/index/evolving-our-structure/ ("We want to be able to operate and get resources in such a way that we can make our services broadly available to all of humanity, which currently requires hundreds of billions of dollars and may eventually require trillions of dollars.").

[45] Kevin Okemwa, *OpenAI Started as a "Countervailing Force" to Google—Did Elon Musk and Sam Altman Torpedo DeepMind's Plans to Dictate AGI?*, Windows Cent. (June 26, 2025), https://www.windowscentral.com/software-apps/openai-started-as-a-countervailing-force-to-google (last visited Feb. 23, 2026).

[46] *OpenAI Inc*, ProPublica: Nonprofit Explorer, https://projects.propublica.org/nonprofits/organizations/810861541 (last visited Oct. 28, 2026). For more on the genesis of OpenAI, see Karen Hao, Empire of AI 46–53 (2025) and Merchant, *supra* note 42, at 9–10.

[47] Tax documents list Elon Musk, Sam Altman and Chris Clark on the board of directors, and Chris Clark and Jonathan Levy as officers. Form 1023, *supra* note 4, Supplemental Responses at 2–3.

[48] Mosaic Ventures, *Mosaic Ventures in conversation with Y Combinator President, Sam Altman*, at 24:14 (YouTube, Mar. 28, 2017), https://www.mosaicventures.com/blog/mosaic-ventures-in-conversation-with-y-combinator-president-sam-altman [https://perma.cc/W5NV-SXTU] (citation modified). *See also* Web Desk, *Why OpenAI Is a Prime Example of the Ethical Limits of Capitalism*, Digit. Info. World (Dec. 9, 2025), https://www.digitalinformationworld.com/2025/12/why-openai-is-a-prime-example-of-the-ethical-limits-of-capitalism.html; Kelsey Piper, *OpenAI's Nonprofit Structure Was Supposed to Protect You. What Went Wrong?*, Vox (Apr. 24, 2025), https://www.vox.com/future-perfect/410261/openai-non-profit-transition-letter-sam-altman-artificial-intelligence.

founders and other organizations amounted to $1 billion,[49] and its application to operate as a tax-exempt organization notes that it "does not plan to play any role in developing commercial products or equipment."[50]

2. Pivot to "Capped-Profit" Structure

In 2019, the company sought more capital to sustain the compute infrastructure needed to scale AI. The conventional scaling paradigm consists in leveraging massive amounts of computation to achieve progress, so AI companies have massively invested in compute to process enormous amounts of data, and perform calculations at scale.[51] In the span of two years alone, OpenAI's computing costs jumped from $2.3 to $30.5 million.[52] To finance this massive expense, it noted on its website that it needed to offer more attractive investment opportunities through a self-styled "capped-profit" business structure, a bespoke investment vehicle that distributed profits up to a predetermined multiple of the initial investment.[53]

Though framed as a transition to a single limited partnership (or L.P., a well-established legal entity explicitly recognized under Delaware law), the resulting corporate structure actually consisted of at least 20 interlocking legal entities evolving over time.[54] Taken together, this fleet of companies formed a joint venture with the non-profit, raising capital through tailored and confidential contractual agreements. Nominally overseen by the original non-profit corporation, several for-profit legal entities offered equity to raise capital investors and attract employees. Figure 6 outlines the main entities involved in the capped-profit business model of OpenAI, with each one detailed below.

---

[49] *Introducing OpenAI*, OpenAI, (2015), https://openai.com/index/introducing-openai/ [https://perma.cc/9JWR-H5C5] (claiming that several named private funders had "[i]n total . . . committed $1 billion"); *but see* Mark Harris, *Elon Musk Used to Say He Put $100M in OpenAI, but Now It's $50M: Here Are the Receipts*, TechCrunch (May 17, 2023), https://techcrunch.com/2023/05/17/elon-musk-used-to-say-he-put-100m-in-openai-but-now-its-50m-here-are-the-receipts/ (noting that actual donations from the company's inception to 2021 only amounted to $133 million); *see also* ProPublica, *OpenAI Inc - Tax Filing - 990* (2021), https://projects.propublica.org/nonprofits/organizations/810861541.

[50] OpenAI, *Internal Revenue Service Form 1023, supra* note 4, at Part IV.

[51] Richard Sutton, *The Bitter Lesson*, Incomplete Ideas (Mar. 13, 2019), http://www.incompleteideas.net/IncIdeas/BitterLesson.html, [https://perma.cc/VB2B-FVXR] (scaling with more compute power yields the most improvements in machine learning); *see also* Jai Vipra & Sarah M. West, *Computational Power and AI*, AI Now Institute 4 (Sep. 27, 2023), https://ainowinstitute.org/publications/compute-and-ai [https://perma.cc/QD3G-WSXE] (defining compute as the combination of chips, graphic processing unit software, domain-specific languages, data management software and data center infrastructure such as connecting cables, servers and cooling equipment); Jared Kaplan et al., *Scaling Laws for Neural Language Models* (Jan. 23, 2020), http://arxiv.org/abs/2001.08361; *but see* Andrew Lohn, *Scaling AI*, Geo. Ctr. for Sec. & Emerging Tech., https://cset.georgetown.edu/publication/scaling-ai/ (last visited Feb. 18, 2025) (calling into question the future of scaling due to diminishing returns).

[52] *OpenAI Inc 2016 Tax Filing*, *supra* note 46; *OpenAI Inc - Tax Filing - 990*, ProPublica (2018), https://projects.propublica.org/nonprofits/organizations/810861541.

[53] Alexandra Andhov, *OpenAI's Transformation: From a Non-Profit to a 157 Billion Valuation*, 46 Bus. L. Rev. 2, 9 (2025), https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\BULA\BULA2025003.pdf (describing OpenAI's "capped-profit" model while noting that "capped-profit" is an entirely novel concept that "lacks legal standing and recognition").

[54] For a more detailed account of OpenAI's corporate structure, see *The OpenAI Files*, https://www.openaifiles.org/restructuring [https://perma.cc/BU5X-PUCM] (last visited July 4, 2025) (including educated guesses about subsidiaries); *see also* Andhov, *supra* note 53, at 8–9 (mapping parent companies and subsidiaries).

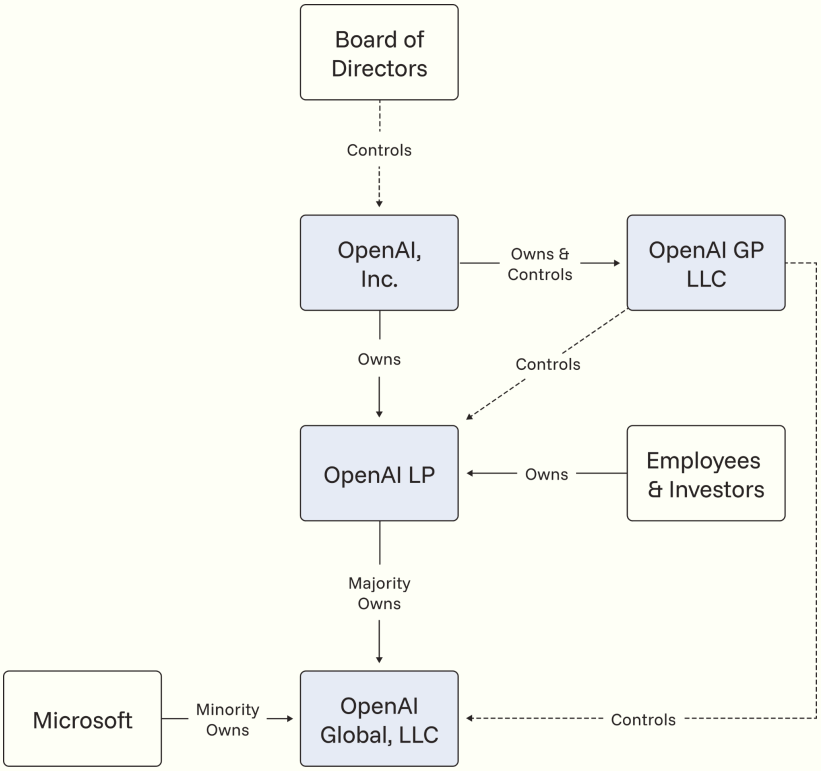


Figure 6: OpenAI Capped-Profit Structure.

Non-Profit OpenAI Inc. Retains Ultimate Control Over Subsidiaries

This is the original non-profit OpenAI legal structure. As a non-stock corporation, it doesn't issue stock to shareholders. Instead, members hold rights similar to shareholders.[55] It was the "mothership" of the OpenAI nexus, technically controlling all other entities. Some board members were also employees whose compensation depended on the company's valuation, but they did not vote on decisions that could trigger conflicts of interest.[56] Under IRS rules for tax-exempt entities, profits belong to the charity.[57] Employee and executive compensation must be at market rates.[58]

OpenAI GP LLC Controls For-Profit Entities

This limited liability company was designed to allow the original non-profit portion of the OpenAI ecosystem to control its various for-profit components. Concretely, the non-profit OpenAI Inc. was the

[55] OpenAI, *Certificate of Incorporation and Bylaws*, *supra* note 40, Art. II (reflecting that the members of OpenAI Inc. are the directors of the corporation).
[56] *Id.* at Art. VII (requiring that only disinterested board members vote on such matters); *see also* Darryll K. Jones, *The OpenAI Corporate Structure*, Nonprofit L. Prof Blog (Jan. 8, 2024), https://web.archive.org/web/20250618175534/https://lawprofessors.typepad.com/nonprofit/2024/01/the-openai-corporate-structure.html (last visited July 7, 2025) (pointing out that some directors' compensation is tied to OpenAI's market performance).
[57] I.R.C. § 501(c)(3).
[58] *Id.* § 4958.

sole member of OpenAI GP LLC, which in turn controlled two for-profit companies: OpenAI LP and OpenAI Global LLC.

Holding Company OpenAI LP Distributes Profits to Early Investors and Employees

OpenAI shifted the majority of its employees from the non-profit to OpenAI LP.[59] As a limited partnership, it consisted of general partners managing operations and limited partners investing in the for-profit company. OpenAI GP LLC was the sole manager, giving the board of non-profit OpenAI Inc. full control over OpenAI LP. In terms of ownership, investors and employees were limited partners of OpenAI LP. The return of early investors was capped at 100x their investment, a rate expected to lower in future funding rounds.[60] Microsoft then invested $1 billion in cash and compute credits on its Azure platform, with the exact return cap never made public. To attract sought-after talent, employees got "profit participation units" (PPUs), which were essentially contractual rights to a share of future profits in lieu of stock in traditional corporations.[61] Employees were offered participation capped at 10x the value at the time of hiring.[62] Employees could sell these PPUs to investors when the company triggered a tender offer, as it did in 2024 at a $157 billion valuation.[63] To give a sense of the magnitude of profits employees stood to make, OpenAI's valuation grew 10x from 2023 to 2025, from $29 to $300 billion. In the late 2024 tender offer, each eligible employee could sell up to $10 million in vested PPUs, with SoftBank committing $1.6 billion to buy from roughly 400 current and former staff; a subsequent October 2025 offer raised the individual cap to $30 million, with approximately 75 employees selling the maximum, and 600 employees collectively liquidating $6.6 billion in PPUs.[64]

The inner workings of the LP remain out of the public eye. An excerpt of the partnership agreement stipulates that the "[p]artnership exists to advance OpenAI Inc's mission of ensuring that safe artificial general intelligence is developed and benefits all of humanity," noting that the mission takes precedence over profit-making obligations.[65] The tax filings of the non-profit arm also indicate that the

[59] *OpenAI LP,* OpenAI, *supra* note 44.
[60] *OpenAI LP,* OpenAI, *supra* note 44, at 5; *but see* Andhov, *supra* note 53, at 9 (noting that the malleable partnership agreement could modify the cap).
[61] Nino Paoli, *Mark Zuckerberg Has a Literal List of AI All-Stars to Whom He's Offering Fantastic Sums of Money. An Expert Explains Why He's Hiring so Aggressively*, Fortune (July 1, 2025), https://fortune.com/2025/07/01/mark-zuckerberg-meta-superintelligence-lab-team-list-openai-edgar-perez/ (last visited July 4, 2025).
[62] Brian Nguyen, *OpenAI PPUs: How OpenAI's Unique Equity Compensation Works*, Levels.fyi Blog (June 19, 2023), https://www.levels.fyi/blog/openai-compensation.html [https://perma.cc/7V4E-DL8U].
[63] *Id.*; *see also* Jack Raines, *OpenAI Isn't Selling Equity—It's Selling Shares of Profits That May Never Come. Investors Can't Get Enough*, Sherwood News (Sep. 10, 2024), https://sherwood.news/business/openai-isnt-selling-equity-its-selling-shares-of-profits-that-may-never-come/ (last visited June 30, 2025); Sherin Shibu, *Some OpenAI Employees Are About to Become Millionaires*, Entrepreneur (Dec. 18, 2024), https://www.entrepreneur.com/business-news/some-openai-employees-are-about-to-become-millionaires/484488.
[64] *Deals - OpenAI - Company Profile*, Pitchbook, https://content.pitchbook.com/share/quick-link/profile/170ba480-ba15-4b9e-96dc-9ff5022e2634?hash=2cc522aca06a7b95f74a376a96ac94ec762545095cda670c5257f355aa7cf312 (last visited July 18, 2025); *New Funding to Build towards AGI*, OpenAI (Mar. 31, 2025), https://openai.com/index/march-funding-updates/ (last visited July 18, 2025); Kyle Wiggers, *OpenAI Raises $6.6B and Is Now Valued at $157B*, TechCrunch (Oct. 2, 2024), https://techcrunch.com/2024/10/02/openai-raises-6-6b-and-is-now-valued-at-157b/; Berber Jin, *How a Job at OpenAI Became the Greatest Lottery Ticket of the AI Boom*, Wall St. J. (May 11, 2026), https://www.wsj.com/tech/openai-employee-stock-sales-71ed10bd; Sharon Goldman, *Hundreds of OpenAI's Current and Ex-Employees Are About to Get a Huge Payday by Cashing Out up to $10 Million Each in a Private Stock Sale*, Fortune (Dec. 17, 2024), https://fortune.com/2024/12/17/hundreds-openai-employees-10-million-payday-softbank-stock-tender-offer-details/.
[65] *OpenAI LP,* OpenAI, *supra* note 44.

partnership agreement prioritizes the non-profit mission over profits,[66] but the precise mechanics of the profit-sharing distribution and decision-making processes remain confidential.[67] Based on confidential documents, Fortune magazine reported that profits would initially flow to investors and employees, then would go towards the non-profit.[68] The LP was a majority owner of the second for-profit company, OpenAI LLC.

For-Profit OpenAI Global LLC in Partnership With Microsoft

In order to finance its massive investment in computation power, OpenAI ramped up its partnership with technology giant Microsoft through this limited liability company. In 2023, Microsoft invested over $10 billion in return for a "minority economic interest" in the company's profits.[69] The non-profit arm retained full control as the sole manager of the LLC (via OpenAI GP LLC), and owned a majority stake via OpenAI LP. The operating agreement of a limited liability company is off-limits to the public, save for excerpts the company voluntarily disclosed. The public excerpt reiterates the above warning that mission overrides profit, cautioning that "it would be wise to view an investment in OpenAI Global, LLC. in the spirit of a donation," as no return on investment is guaranteed.[70] That said, the governance crisis detailed below suggests that profit-oriented actors had significant sway despite lacking a formal say in the company's decision-making process.

3. Governance Challenges with Capped-Profit Structure

The bespoke capped-profit structure sought to ensure that the non-profit board retained total control of all the for-profit entities to safeguard the mission over profits.[71] Yet power struggles over the leadership of OpenAI indicate a disconnect between formal and de facto control, with profit-driven stakeholders like Microsoft holding the lion's share of power despite no official vote.[72] The failed ouster

[66] OpenAI Inc - Tax Filing - 990, at 49, ProPublica (2021), https://projects.propublica.org/nonprofits/organizations/810861541/202243199349314989/full.
[67] Del. Code Ann. tit. 6, § 17-201 (2024) (requiring the publicly filed certificate of limited partnership to set forth only the partnership's name, registered office and agent, and general partners, not the partnership agreement); *see also OpenAI LP*, OpenAI, *supra* note 44 (disclosing only the excerpt that opens investor and employee paperwork).
[68] Jessica Mathews & Jeremy Kahn, *Inside the Structure of OpenAI's Looming New Investment from Microsoft and VCs*, Fortune, Jan. 11, 2023, https://fortune.com/2023/01/11/structure-openai-investment-microsoft/ [https://perma.cc/FV45-WNJV] (last visited Nov. 25, 2024).
[69] *Our Structure*, OpenAI, https://openai.com/our-structure/ (last visited Feb. 9, 2025) [https://web.archive.org/web/20250210080054/https://openai.com/our-structure/]; Alex Hern & Lisa O'Carroll, *Microsoft's Investment in OpenAI May Face EU Scrutiny, Officials Say*, Guardian (Jan. 9, 2024), https://www.theguardian.com/technology/2024/jan/09/microsoft-investment-openai-chatgpt-european-union-eu-merger-investigation (noting that OpenAI previously described Microsoft as a minority owner but changed the description as merger scrutiny increased). For details about the profit-distribution arrangements, see Kahn, *supra* note 23. The cap on Microsoft's January 2023 $10 billion investment is not public, but reporting suggests a fork between 20x and "single digits." Hao, *supra* note 46 (reporting a 20x cap on $1 billion investment in June 2019); The Ezra Klein Show, *Sam Altman on the A.I. Revolution, Trillionaires and the Future of Political Power*, (N.Y. Times, June 11, 2021), https://www.nytimes.com/2021/06/11/opinion/ezra-klein-podcast-sam-altman.html (reporting a single digit cap in June 2021); Matt Levine et al., *Who Owns OpenAI?*, Bloomberg (Oct. 21, 2024), https://www.bloomberg.com/opinion/articles/2024-10-21/who-owns-openai (breaking down the profit distribution among various investors).
[70] *Our Structure*, OpenAI, *supra* note 69.
[71] Ellen P. Aprill, Rose C. Loui & Jill R. Horwitz, *Board Control of a Charity's Subsidiaries: The Saga of OpenAI*, 182 Tax Notes Fed. 289, 292–93 (2024), https://papers.ssrn.com/abstract=4720202.
[72] Frederik H. Pedersen, *Ownership at OpenAI: From the Perspective of Enterprise Foundation Governance* (Apr. 15, 2024) (unpublished manuscript), https://papers.ssrn.com/abstract=4795279.

of CEO Sam Altman suggests that de facto pressure from employees and investors overrode the board's formal control. In November 2023, the board fired Altman on the grounds that he "was not consistently candid in his communications."[73] Though the full story remains murky,[74] the role of powerful constituencies to override the board appears quite clear. The board reversed course, in no small part due to employee and investor pressure. Not only was Altman reinstated, but the board members who had attempted to fire him resigned, flipping the governance structure on its head. The influence of employees and investors points to a disconnect between the governance of OpenAI on paper and the gravitational pull of other actors in reality. These "superstakeholders"[75] lack a formal vote on the board, but hold formidable sway over the company. The two sections below explain the leverage and influence of employees and investors over the company's decisions.

Employees Mobilizing for Altman

Employees contributed to Altman's reinstatement by threatening to quit en masse in protest of his firing.[76] Without its highly sought-after specialized workforce, OpenAI would have come to a halt. Employees thus had significant leverage when they rallied around Altman. Perhaps not coincidentally, Altman's firing occurred as the company was gearing up for a tender offer. This opportunity for employees to cash out on their PPUs was tied to the company's valuation, and the upheaval in OpenAI's leadership threatened to erase the total $86 billion valuation—more than a 50% jump in a matter of a few months.[77] This created a powerful incentive for employees to put their weight behind Altman to restore continuity and preserve the value of their equity—typically more than half their compensation.[78]

---

[73] *OpenAI Announces Leadership Transition*, OpenAI (Nov. 17, 2023), https://openai.com/index/openai-announces-leadership-transition/. For a deeper analysis of the factors leading to Altman's firing, including safety concerns with rapid commercialization, see Hao, *supra* note 46, at 355-76; Olson, *supra* note 13, at 269-79.

[74] *Review completed & Altman, Brockman to continue to lead OpenAI*, OpenAI (Mar. 8, 2024), https://openai.com/index/review-completed-altman-brockman-to-continue-to-lead-openai/ (claiming that Altman's behavior "did not mandate removal," which instead resulted from "a breakdown in trust between the prior Board and Mr. Altman"); Hayden Field, *Former OpenAI Board Member Explains Why CEO Sam Altman Got Fired before He Was Rehired*, CNBC (May 29, 2024), https://www.cnbc.com/2024/05/29/former-openai-board-member-explains-why-ceo-sam-altman-was-fired.html (alleging that Altman withheld information, made material misrepresentations, and outright lied to the board).

[75] *Developments in the Law—Amoral Drifts in Corporate AI Governance*, *supra* note 10, at 1653. *See also* Pedersen, *supra* note 72; Aprill, Loui & Horwitz, *supra* note 71 (both noting the *de facto* pull of Altman and equity holders).

[76] Jacob Kastrenakes, *Hundreds of OpenAI Employees Threaten to Resign and Join Microsoft*, Verge (Nov. 20, 2023), https://www.theverge.com/2023/11/20/23968988/openai-employees-resignation-letter-microsoft-sam-altman; Cade Metz et al., *Sam Altman Is Reinstated as OpenAI's Chief Executive*, N.Y. Times (Nov. 22, 2023), https://www.nytimes.com/2023/11/22/technology/openai-sam-altman-returns.html.

[77] PitchBook Profile - OpenAI, *supra* note 6 (reporting OpenAI's $29 billion valuation in April 2023 and $86 billion valuation in January 2024); Reuters, *OpenAI's $86 Billion Share Sale in Jeopardy after Altman Firing, The Information Reports*, Reuters (Nov. 18, 2023), https://www.reuters.com/technology/openais-86-bln-share-sale-jeopardy-following-altman-firing-information-2023-11-18/.

[78] Paolo Confino, *OpenAI's Meltdown: What to Know about Sam Altman's Ouster, the Employee Revolt, and Microsoft's Plot Twist*, Fortune, Nov. 20, 2023, https://fortune.com/2023/11/20/openai-sam-altman-microsoft-ai-what-happened/ (last visited July 4, 2025); Reuters, *supra* note 77; Nguyen, *supra* note 62 (describing an OpenAI offer for a Member of Technical Staff position stipulating base salary of $300,000 and equity worth of $800,000 with a 10x profit cap).

Major Investors

As of Altman's firing, Microsoft had committed $13 billion to OpenAI.[79] The leading investor at the time, it carried significant weight despite lacking a voice in OpenAI's formal governance.[80] As a publicly listed company whose own valuation was tied to the lab's performance, Microsoft presumably prioritized a leadership that would align with its vision to ship profitable products. When the sudden ouster of Altman threatened to throw OpenAI into disarray, Microsoft could have withheld billions of undisbursed investment and computing infrastructure critical to OpenAI's activities. It is also deeply intertwined in OpenAI's supply chain, leading its CEO to note that "We are in there. We are below them, above them, around them. We do the kernel optimizations, we build tools, we build the infrastructure."[81] In addition to providing OpenAI with essential infrastructure, Microsoft had licensing rights giving it preferential access to OpenAI technology.[82] Investor and employee pressure also compounded, as Microsoft's standing offer to absorb any OpenAI employee who resigned gave the threat of a mass walkout real teeth.[83]

Although the non-profit board technically controlled every for-profit entity downstream, the outcome of its power struggle to fire Altman indicates that investors and employees actually called the shots. As nonprofit consumer advocates noted, it seemed as if the for-profit component effectively controlled the non-profit component of OpenAI.[84] This calls into question the board's ability to put the public interest mission above the profit-aligned interests of these superstakeholders.[85] Indeed, the pressure to make profitable products strained the capped-profit model, so much so that the company considered abandoning its non-profit vocation altogether. The company faced pressure in opposite directions: profit-oriented stakeholders wanted a switch to a for-profit model[86] that would have removed the

---

[79] Jordan Novet, *Microsoft's $13 Billion Bet on OpenAI Carries Huge Potential along with Plenty of Uncertainty*, CNBC (Apr. 8, 2023), https://www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html.

[80] Max Chafkin & Dina Bass, *Microsoft's Sudden AI Dominance Is Scrambling Silicon Valley's Power Structure*, Bloomberg (June 15, 2023), https://www.bloomberg.com/news/features/2023-06-15/microsoft-prepares-to-cash-in-on-openai-partnership-with-copilot (describing a 30% increase in Microsoft's share value after the success of ChatGPT).

[81] Intelligencer Staff, *Satya Nadella on Hiring the Most Powerful Man in AI*, Intelligencer (Nov. 21, 2023), https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-on-hiring-sam-altman.html.

[82] Microsoft Corporate Blogs, *Microsoft and OpenAI Evolve Partnership to Drive the next Phase of AI*, The Official Microsoft Blog (Jan. 21, 2025), https://blogs.microsoft.com/blog/2025/01/21/microsoft-and-openai-evolve-partnership-to-drive-the-next-phase-of-ai/ (describing Microsoft's exclusive access to the OpenAI API for Azure and rights to OpenAI IP for Copilot).

[83] Tom Gerken, *Microsoft Offers to Match Pay of All OpenAI Staff*, BBC (Nov. 21, 2023), https://www.bbc.com/news/technology-67484455; *but see* Kali Hays, Ashley Rafieyan & Darius Stewart, *OpenAI Employees Really, Really Did Not Want to Go Work for Microsoft*, Bus. Insider (Dec. 6, 2023), https://www.businessinsider.com/openai-employees-did-not-want-to-work-for-microsoft-2023-12 (last visited July 7, 2025) (reporting that employees were privately dreading leaving an agile startup for the "biggest and slowest" big tech company).

[84] Robert Weissman, *Letter to California Attorney General on OpenAI's Nonprofit Status*, Public Citizen (Jan. 9, 2024), https://www.citizen.org/article/letter-to-california-attorney-general-on-openais-nonprofit-status/ (calling into question effective non-profit control after the board's ouster).

[85] Tim Bradshaw et al., *How Microsoft's Multibillion-Dollar Alliance with OpenAI Really Works*, Fin. Times (Dec. 15, 2023), https://www.ft.com/content/458b162d-c97a-4464-8afc-72d65afb28ed (describing Microsoft's pressure to ship products as it "really wanted to be deploying AI products and be an AI leader and make money"); Robert Weissman, *supra* note 84, at 3–4 (describing a consensus view that for-profit interests overtook the non-profit interests).

[86] Hayden Field & Kate Rooney, *OpenAI Funding Round Could Be Cut by $10 Billion If For-Profit Conversion Doesn't Occur by Year-End*, CNBC (Mar. 31, 2025), https://www.cnbc.com/2025/03/31/openai-funding-could-be-cut-by-10-billion-if-for-profit-move-lags.html (reporting that SoftBank's portion of an OpenAI investment round would decrease by $10 billion if OpenAI remained a non-profit).

non-profit from its controlling position,[87] whereas some early investors, former employees, competitors and observers opposed departures from the non-profit commitments.[88] As a middle ground, the company has pivoted to a public benefit corporation that maintains the controlling position of the non-profit but allows it to make profits.

4. Public Benefit Corporation Reorganization

OpenAI restructured as a public benefit corporation (PBC) in October 2025, with the mission to "ensure that artificial general intelligence benefits all of humanity, including by conducting and/or funding artificial intelligence research."[89] The original non-profit, now called the OpenAI Foundation, holds 26% of stock in the public benefit corporation, leaving Microsoft with 27% and employees and other investors with the remaining 47%.[90] Funded by its shares in the PBC, the Foundation focuses on philanthropy.[91] In addition to its current 26% stake, the Foundation stands to gain additional shares if share prices grow tenfold over 15 years.[92] The Foundation's financial interest in the PBC fundamentally misaligns the governance structure, as its own stake in the company's share value undermines its ability to prioritize the mission over profits. Given that the new structure removed the profit caps that guaranteed the non-profit would eventually reap 100% of the profits, some have also construed the reorganization as shortchanging the non-profit.[93]

---

[87] *Why OpenAI's Structure Must Evolve to Advance Our Mission*, OpenAI (Dec. 27, 2024), https://openai.com/index/why-our-structure-must-evolve-to-advance-our-mission/ (last visited June 27, 2025) (proposing to split the company into distinct entities, with a non-profit and a public benefit corporation directly responsive to investors); *see also* Rose C. Loui, *Can OpenAI Abandon Its Non-Profit "Purpose"?*, ProMarket (Apr. 23, 2025), https://www.promarket.org/2025/04/23/can-openai-abandon-its-non-profit-purpose/.

[88] *Musk v. Altman,* No. 4:24-cv-04722 (N.D. Cal. 2026) (alleging unjust enrichment based on false claims that OpenAI was a safety-oriented non-profit organization); *Not For Private Gain* (Apr. 12, 2025), https://notforprivategain.org (former employees and corporate governance experts expressing concerns about safeguarding the mission under a for-profit model); Alex Heath, *Meta Asks the Government to Block OpenAI's Switch to a for-Profit*, Verge (Dec. 13, 2024), https://www.theverge.com/2024/12/13/24320880/meta-california-ag-letter-openai-non-profit-elon-musk (reporting on Meta's claims that "if OpenAI's new business model is valid, non-profit investors would get the same for-profit upside as those who invest the conventional way in for-profit companies while also benefiting from tax write-offs bestowed by the government and, ultimately, the public," and that "[t]hat would distort the market by essentially requiring any startup seeking to remain competitive to adopt the same playbook."); Loui, *supra* note 87; Krystal Hu et al., *OpenAI Dials Back Conversion Plan, Nonprofit to Retain Control*, Reuters (May 5, 2025), https://www.reuters.com/business/openai-remain-under-non-profit-control-change-restructuring-plans-2025-05-05/ (last visited June 25, 2025); *An Open Letter to OpenAI*, https://www.openai-transparency.org/ (last visited Sep. 23, 2025); Letter to California Attorney General on OpenAI's Nonprofit Status, *supra* note 84; *Coalition for AI Nonprofit Integrity*, https://ab501.com/[https://web.archive.org/web/20250416225539/https://ab501.com/] (last visited Sep. 4, 2025).

[89] *Certificate of Incorporation of OpenAI Group PBC*, art. III (2025).

[90] *Our Structure,* OpenAI, *supra* note 69.

[91] *Evolving OpenAI's Structure*, OpenAI, *supra* note 44 (aspiring to set up "the largest and most effective nonprofit in history that will be focused on using AI to enable the highest-leverage outcomes for people"); *see also OpenAI Board of Directors Announces Advisors to Nonprofit Commission*, https://openai.com/index/nonprofit-commission-advisors/ (last visited July 8, 2025) (setting up a commission to inform philanthropic efforts); *Our Structure*, OpenAI, *supra* note 69.

[92] *Our Structure*, OpenAI, *supra* note 69.

[93] The Midas Project, *The Midas Project Statement on OpenAI's Restructuring* (Oct. 28, 2025), https://www.themidasproject.com/articles/the-midas-project-statement-on-openais-restructuring (last visited Nov. 27, 2025).

Control-wise, the Foundation appoints the board of the PBC.[94] All but one of the directors of the Foundation are also on the PBC board, and within a year two directors will serve exclusively on the Foundation's board.[95] With the PBC board essentially a subset of the Foundation, there are concerns that the non-profit's only actual power is to fire itself.[96] Furthermore, the other profit-oriented shareholders retain the capacity to fire a third of the PBC's board.[97]

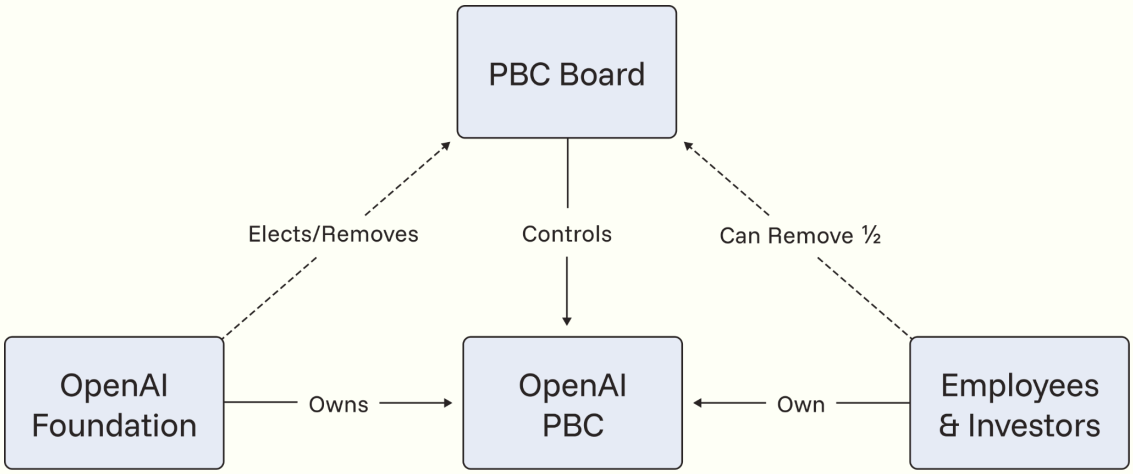


Figure 7: OpenAI's Public Benefit Corporation restructuring.

Despite some safeguards, the PBC restructuring weakens OpenAI's already brittle governance. The attorneys general of Delaware and California extracted some commitments to the non-profit mission in exchange for withholding their objections to the reorganization.[98] A Safety and Security Committee has to remain under the control of the Foundation and retains a right of approval over safety and security matters, including the ability to halt the release of models and systems.[99] The PBC's board also has to

[94] *Certificate of Incorporation of OpenAI Group PBC*, *supra* note 89, art. IV.2(b)(i) (describing the voting power of the Class N common stock shares owned by the Foundation); Ann M. O'Leary & William Savitt, *Response to Petition for Attorney General Action To Protect The Charitable Nature Of OpenAI's Assets*, 7 (2025); *Evolving OpenAI's Structure*, OpenAI, *supra* note 44 (noting that "[w]e made the decision for the nonprofit to retain control of OpenAI after hearing from civic leaders and engaging in constructive dialogue with the offices of the Attorney General of Delaware and the Attorney General of California").
[95] *Our Structure*, OpenAI, *supra* note 69.
[96] The Midas Project, *supra* note 93.
[97] *Certificate of Incorporation of OpenAI Group PBC*, *supra* note 89, art. V.3 (stating a maximum number of director removals for a 12-month period by shareholders other than the board and requiring a 2/3 majority to proceed).
[98] Khari Johnson & Levi Sumagaysay, *OpenAI Just Cut a Deal with California. Critics Say It's Full of Holes*, CalMatters (Oct. 30, 2025), https://calmatters.org/economy/technology/2025/10/openai-restructuring-deal-full-of-holes-critics-say/; Chase Difeliciantonio & Christine Mui, *OpenAI Gets Green Light from California for Multibillion-Dollar Makeover, but Still Faces Hurdles*, Politico (Oct. 28, 2025), https://www.politico.com/news/2025/10/28/openai-business-restructuring-california-00625383; Matthew Marshall, *AG Jennings Completes Review of OpenAI Recapitalization*, State Del. News (Oct. 28, 2025), https://news.delaware.gov/2025/10/28/ag-jennings-completes-review-of-openai-recapitalization/.
[99] Rob Bonta, Att'y Gen. of Cal. & OpenAI, Inc., Memorandum of Understanding, ¶¶ 9–11 (Oct. 27, 2025) [hereinafter CA MOU]; Letter from Kathleen Jennings, Att'y Gen. of Del., to Che Chang, Gen. Counsel, OpenAI, Inc., re: OpenAI, Inc.'s Corporate Restructuring ¶¶ 7-9 (Oct. 28, 2025) [hereinafter DE MOU].

focus on the mission in matters of safety and security, to the exclusion of the financial interest of shareholders.[100] However, these commitments may not work well in practice. For context, the former structure had even stronger incentives to prioritize the mission over profits: assets locked in a non-profit and no shareholders pressing for profit. And even then, it could not carry out the firing of the CEO, let alone prevent the PBC restructuring which introduced profit-driven shareholders in the governance structure.

More broadly, the PBC structure is itself limited when it comes to enforcing the mission. OpenAI frames it as a "purpose-driven company structure that has to consider the interests of both shareholders and the mission."[101] But under Delaware law, the directors of a public benefit corporation have significant leeway in balancing shareholder financial interests with the best interests of those materially affected by the corporation's conduct, and the specific public benefit identified in the Certificate of Incorporation—in this case funding AI research to benefit the public. As long as their decisions are informed, disinterested and "not such that no person of ordinary, sound judgment would approve," directors will steer clear of liability.[102] Under the business judgment rule, courts will not second-guess a board's balancing decision as long as it has a rational basis. Because the PBC legal regime is fairly new and hard to challenge, there is relatively little guidance as to how directors should balance profit, mission and stakeholder interests. Furthermore, only directors and certain shareholders can challenge company decisions. The section below, "Limited Director Liability for Prioritizing Profit Over Mission," explains in greater detail the potential and limitations of the PBC regime for prioritizing profits over the mission. In sum, the absolute primacy of the mission hardcoded in the original non-profit and in the previous joint venture agreements[103] is now just one among several factors driving OpenAI, and enforcement avenues remain limited.[104]

Changes in corporate structure reflect OpenAI's evolution from a non-profit research lab to a product-driven venture. Having streamlined its uniquely complex web of interlocking legal entities into a single public benefit corporation, today it stands at the crossroads.

---

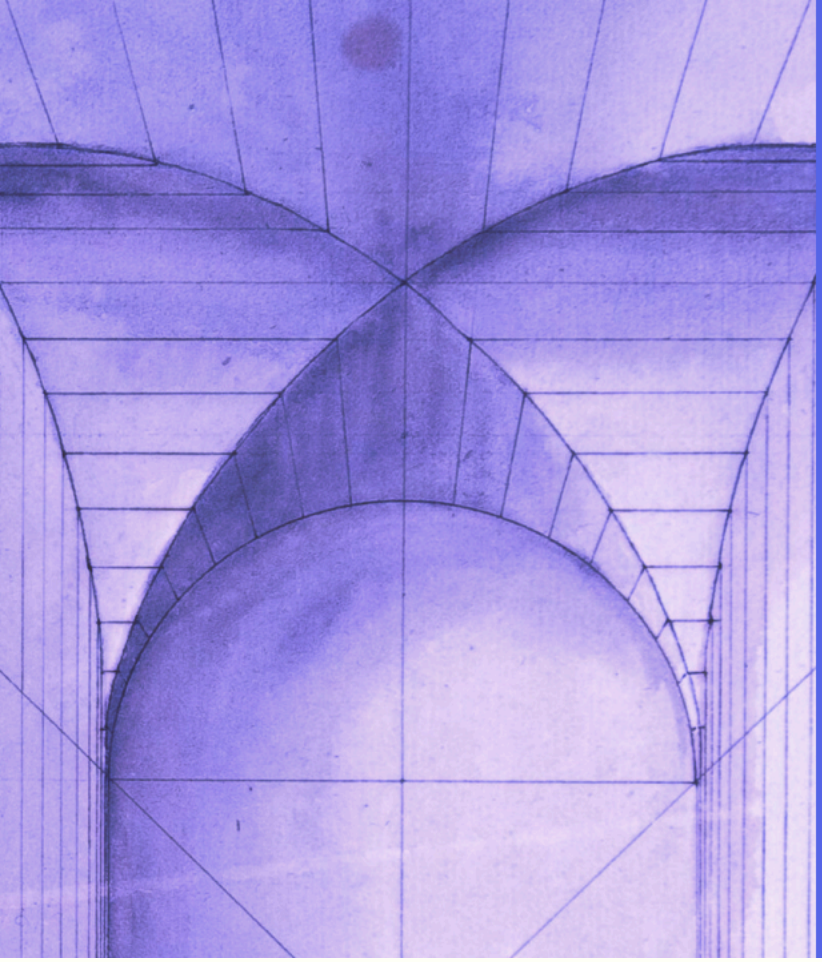
[100] CA MOU, *supra* note 99, ¶ 8; DE MOU, *supra* note 99, ¶ 6; *Certificate of Incorporation of OpenAI Group PBC*, *supra* note 89, art. III ("[T]he Board of Directors of the Corporation (the 'Board of Directors') must solely consider the Mission (and may not consider the pecuniary interests of stockholders or any other interest) in respect of safety and security issues related to the OpenAI enterprise [. . .].").

[101] *Evolving OpenAI's Structure*, OpenAI, *supra* note 44.

[102] Del. Code Ann. tit. 8, § 365, *see also* David Nows & Jeff Thomas, *Delaware's Public Benefit Corporation: The Traditional VC-Backed Company's Mission Driven Twin* (unpublished manuscript) (Sep. 3, 2019), https://papers.ssrn.com/abstract=3447351; *Developments in the Law*, Harv. L. Rev., *supra* note 10, at 1644; *see also The Delaware Way: Deference to the Business Judgment of Directors Who Act Loyally and Carefully*, Del. Corp. L., https://corplaw.delaware.gov/delaware-way-business-judgment/ (last visited July 7, 2025) (describing the business judgment rule, which shields directors from liability so long as they act in good faith, with due care, and without conflicts of interests).

[103] *See supra* notes 65–70 and accompanying text on the partnership agreement of OpenAI LP and the operating agreement of OpenAI Global LLC.

[104] *Update: Not For Private Gain*, *supra* note 88 (denouncing the dilution of the non-profit mission in the new structure).



Anthropic

As with OpenAI, Anthropic has an innovative, bespoke corporate structure. Trustees oversee a public benefit corporation—though the structure is experimental and shareholders can be reversed.

1. Public Benefit Corporation

Anthropic markets its chatbot Claude to the public and to businesses. Just as OpenAI was founded in reaction to Google, Anthropic was founded in reaction to OpenAI. Former employees, including the former VP of research at OpenAI, Dario Amodei, quit over safety and commercialization concerns and set out to create their own company.[105] Founded in 2021, Anthropic presents itself as a safety-oriented company that strives to build "reliable, interpretable, and steerable AI systems."[106] Incorporated as a Delaware public benefit corporation (the same legal vehicle OpenAI recently adopted), its board must balance the same three factors in overseeing the company: the pecuniary interests of stockholders, the best interests of those materially affected by the company's conduct, and the company's mission to "responsibly develop and maintain advanced AI for the long-term benefit of humanity."[107]

---

[105] Olson, *supra* note 13, at 178–82 (detailing the origin story of Anthropic and noting business opportunities as an additional motivating factor).
[106] *Company \ Anthropic*, https://www.anthropic.com/company (last visited July 3, 2025).
[107] Anthropic Certificates of Incorporation, *supra* note 3, at III. The company has tweaked its Certificate of Incorporation twice a year since its incorporation in Jan. 26, 2021, but has consistently maintained this mission since the second iteration of its Certificate of Incorporation in May 24, 2021. Unless otherwise noted, the analysis is based on the latest version available at the time of writing, dated Jan. 27, 2025.

2. Long-Term Benefit Trust

Like OpenAI, Anthropic has been experimenting with its corporate structure. In 2023, it added a long-term benefit trust to oversee its activities.[108] Disinterested trustees incentivize the company to take the long view on safety over rapid commercialization.[109] The trustees control a class of no-value shares slated to elect a majority of the PBC's board of directors over time, with the ability to remove those directors without cause—see Figure 8.[110] As of November 2026, one of the five seats are vacant, with no announced successors. One trustee left to become Chief Global Affairs Officer, thus moving from disinterested oversight to an operational role.

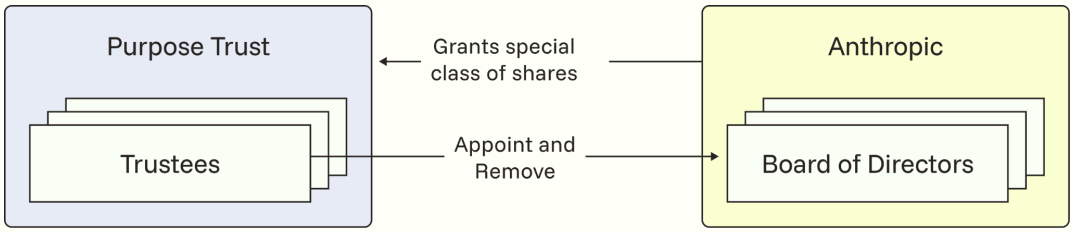


Figure 8: Source: Schuett et al.[111]

The long-term benefit trust is set to play a prominent oversight role via its control of the PBC's board, but several escape hatches constrain it. For example, a supermajority of shareholders can rewrite the rules applicable to the trust's powers—see Figure 9.[112] The architects of Anthropic's structure contend that this flexibility is necessary to ensure its durability.[113]

---

[108] *Anthropic Certificate of Incorporation*, Google Docs IV(D) (May 20, 2024), https://drive.google.com/file/d/17szwAHptolxaQcmrSZL_uuYn5p-N6pex/edit (creating Class T stocks attributed to the trust); *The Long-Term Benefit Trust*, Anthropic, https://www.anthropic.com/index/the-long-term-benefit-trust (last visited Nov. 23, 2023); Amy Simmerman, David Berger & John Morley, *Anthropic Long-Term Benefit Trust*, Harv. L. Sch. F. on Corp. Governance (Oct. 28, 2023), https://corpgov.law.harvard.edu/2023/10/28/anthropic-long-term-benefit-trust/; assisted by the same experts who advised Meta on the Oversight Board, *see Ethical Compass Advisors*, https://www.ethical-compass-advisors.com/ (last visited July 8, 2025).

[109] *The Long-Term Benefit Trust*, Anthropic, *supra* note 108 ("[T]he LTBT can ensure that the organizational leadership is incentivized to carefully evaluate future models for catastrophic risks or ensure they have nation-state level security, rather than prioritizing being the first to market above all other objectives.").

[110] Anthropic Certificate of Incorporation, *supra* note 3, §§ IV (D) (1), VIII (B).

[111] Jonas Schuett, Ann-Katrin Reuel & Alexis Carlier, *How to Design an AI Ethics Board*, 5 AI & Ethics 863, 869 (2024).

[112] Anthropic Certificate of Incorporation, *supra* note 3. at XI (requiring a supermajority to amend the provisions pertaining to the stocks of the trust). *See also* art. IV (D)(4)(iv) (setting a gradually higher threshold for the stockholder supermajority and noting that "the final phase will require 75% of common stock, 50% of Series A Preferred stock as well as 75% of preferred stock, or 85% of the total voting stock."); *see also* Simmerman, Berger, & Morley, *supra* note 108 (outlining three avenues to alter the trust: "by consent of the Voting Trustees and the company's stockholders; by consent of the Voting Trustees and the company's directors prior to the time the Voting Trustees gain the power to elect a majority of the directors; or by a supermajority of stockholders").

[113] Simmerman, Berger, & Morley, *supra* note 108.

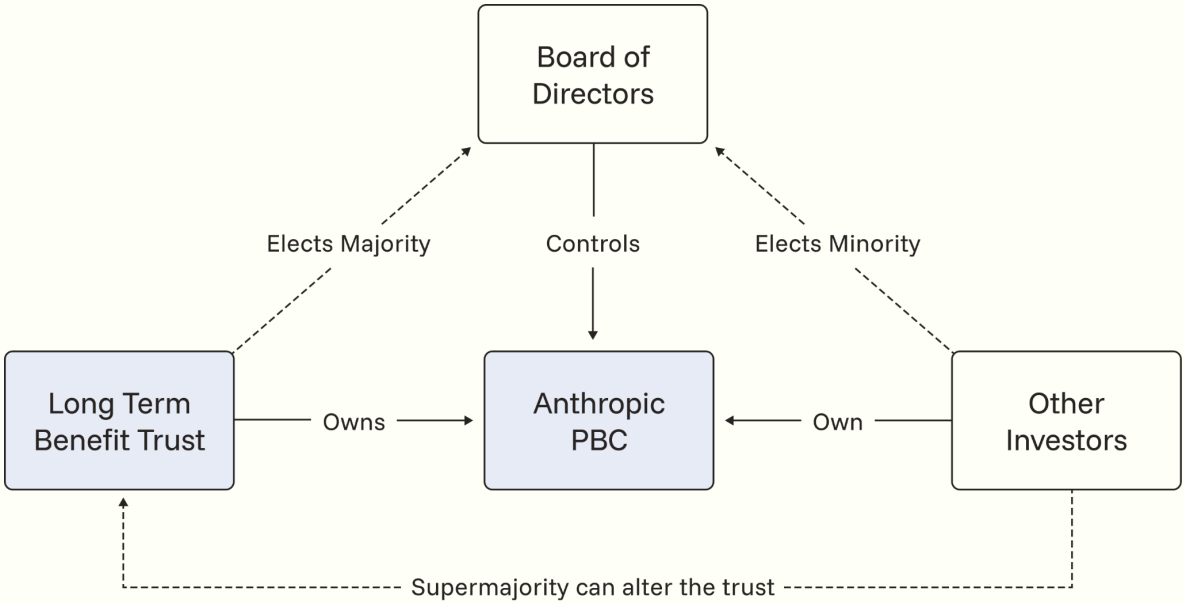


Figure 9: Anthropic's organizational chart.

The trust agreement remains confidential such that its precise mechanics cannot be ascertained. However, a public summary outlines its obligations to ensure that "the company combines its pursuit of profit with the achievement of the company's mission to develop and maintain advanced AI for the long-term benefit of humanity as well as a commitment to the interests of those materially affected by the company's conduct," essentially mirroring the default obligations of a public benefit corporation under Delaware law.[114]

The summary also mentions that the trust is a 'purpose trust' lacking beneficiaries.[115] This is unusual because trusts are typically tripartite arrangements comprising (1) a settlor dedicating assets to the trust; (2) trustees; and (3) beneficiaries. Beneficiaries typically ensure trustees act in good faith, with care and loyalty to prioritize the needs and interests of the beneficiaries.[116] In the absence of beneficiaries, Delaware law provides two main ways to enforce a purpose trust: either the trust agreement appoints enforcers or courts do so.[117] Anthropic went for the first option, appointing the company and shareholders with sufficient equity and tenure to enforce the trust.[118] This is a crucial weakness because a company and its shareholders typically care more about short-term profitability

[114] *Id.;* The Long-Term Benefit Trust, *supra* note 108 (explaining that the trust has the same purposes as Anthropic: "responsibly balances the financial interests of stockholders with the interests of those affected by Anthropic's conduct and our public benefit purpose").
[115] Simmerman, Berger, & Morley, *supra* note 108 see also Del. Code Ann. tit. 12, § 3556 describing purpose trusts.
[116] For a full list of a trustee's fiduciary duties, *see* Mark L. Ascher, Austin W. Scott, & William F. Fratcher, Scott and Ascher on Trusts 17 (6th ed. 2023).
[117] Del. Code Ann. tit. 8, § 3556 (c). As an additional layer of protection, somebody with an interest in the trust can ask to court to be appointed; Adam J. Hirsch, *Delaware Unifies the Law of Charitable and Noncharitable Purpose Trusts*, Estate Planning, 5 (2009), https://papers.ssrn.com/abstract=1481977.
[118] Simmerman, Berger, & Morley, *supra* note 108. Here, "the company" presumably means its Board of Directors, but since the trust agreement is private details cannot be confirmed.

than the longer timeline. Had Anthropic appointed disinterested civil society third parties to enforce the trust, oversight would have been much more robust. As it stands, the trust enforcement structure actually reinforces the very profit-centered pressure trustees are supposed to offset. It deters trustees from ignoring shareholders' interests, but does little to incentivize them to thoroughly monitor the board's duty to consider the mission and the interests of those materially affected by Anthropic's activities.

3. Control and Governance Post-IPO

As this report goes to press, Anthropic has released further information about its structure. In preparation for its initial public through which the company will begin trading shares on the public market, it filed an amended registration statement with the Securities and Exchange Commission on November 2, 2026. The filing indicates that public investors will have limited influence over who sits on the board, but some say over changes to the Trust. The board is structured as follows:

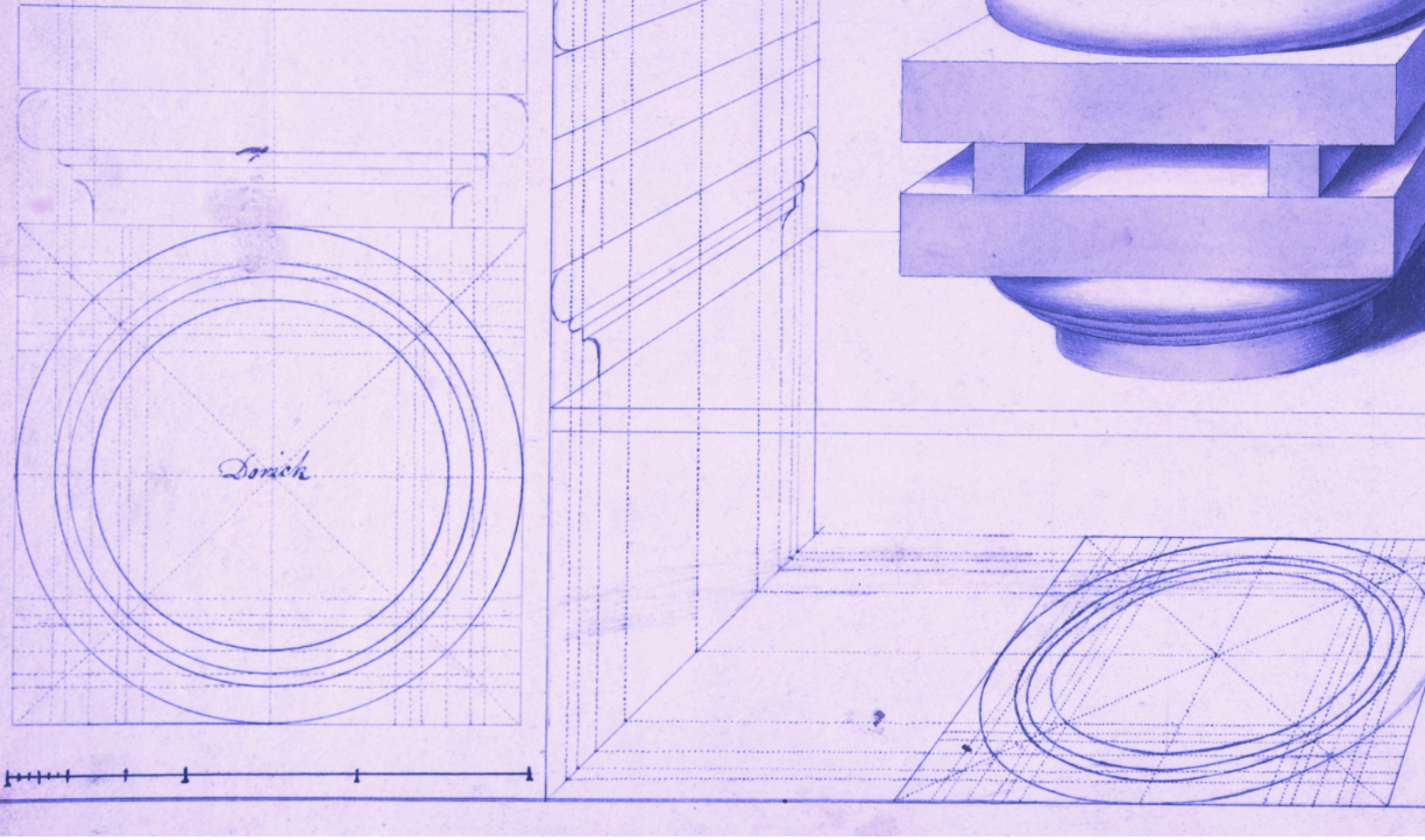
- **3 directors** elected by investors and founders, with the founders holding a majority of the vote
- **3 directors** elected by the Trust
- **1 director** elected by the Trust, subject to approval by a board committee on which directors elected by investors and founders hold two of three seats. If the committee does not approve, employees vote on the Trust's choice.

The founders hold their majority vote through a single special share, held by a limited liability company (LLC) whose members are Anthropic's seven co-founders. That share lapses once fewer than three founders (or their successors) remain members of that company, after a transition period. Because the founders control who fills the seats elected by investors and founders, they have indirect influence over the Trust's fourth seat, the one that gives the Trust its board majority. If the committee and the Trust disagree, employees settle the matter through a special class of voting shares. Employees receive these shares based on their length of service, and receive more for pledging some of their equity to charity, giving up financial value in exchange for a greater say.

The filing also sets out how the Trust can be changed. With the support of 80% of trustees and 80% of the board, its terms can be amended, or its power to elect directors removed altogether. Shareholders can also change the Trust over the trustees' objections. As long as the founders' share exists, this requires the agreement of founders, employees and investors. Once the founder's share lapses, investors alone can do so, with 85% of investor shares.

The diverse corporate models of the three leading research labs each attempt to weave public interest missions into their activities. Whether it is DeepMind's ill-fated Global Interest Company plan, OpenAI's "capped profit" joint venture or Anthropic's long-term benefit trust overseeing a public benefit corporation, all have at some point tried to build corporate structures better aligned with public

interest. The following case studies evaluate how commitments to place the mission above profits play out in practice.



Part II. Business Structures in Action: Assessing Corporate Practices

How do AI companies implement their mission to “benefit humanity”? This section evaluates whether some corporate structures are more conducive to walking the talk on public interest missions by documenting concrete actions in the context of different corporate forms. It finds that overall, there is no meaningful correlation between corporate structure and how companies implement their mission.

The position of AI companies with regard to proposed regulation gives a sense of their willingness to actually agree to external constraints with compliance costs. While they clamor for regulation in the abstract, they tend to oppose specific regulatory proposals. As for self-governance practices, publicly available information suggests that companies use some self-restraint to put safety above profits, though no independent verification allows a definitive determination on the matter. Finally, initial limitations on AI weapons and military applications gradually eroded across most AI companies, each abandoning earlier restraints. This speaks volumes as to their actual commitment to put principles above profit.

As Annexes 1–3 show, corporate form is not a stable trait but a moving target, with each company restructuring at its own pace.[119] Different market positions and maturity levels further confound a definitive cross-company comparison; the more telling evidence often lies within a single company, where conduct can be tracked against its own structural shifts. Finally, the case studies are necessarily time-bound, with events unfolding almost in real time. Given the dynamism of this space, the sample provides a retrospective assessment of company behavior, not a comprehensive up-to-date analysis. Where relevant, developments subsequent to the February 2026 cutoff date are flagged for future research.

[119] Google and OpenAI were corporations and a capped-profit structure throughout the case studies, but Anthropic added the long-term benefit trust after some occurrences. *See* Annex 1 for detailed timelines.



AI Regulation

This section assesses the regulatory policy of AI companies as an indicator of their willingness to promote their public interest mission over their private interest. At the outset, it is important to acknowledge that a pro-regulatory stance is far from a perfect proxy for promoting public interest. First, not all regulation is made equal. Companies may choose to support light-touch regulation to deflect calls for stronger laws. Poorly designed regulation can also be counter-productive, actually securing the discretion and upper hand of companies under a veneer of public legitimacy.[120] With these caveats in mind, regulatory policy remains a good indicator of how companies align mission and action. It conveys the extent to which they are willing to carry out their mission by complying with guardrails that purport to safeguard human well-being against AI's downsides—even if these guardrails are imperfect. In theory, companies purporting to benefit humanity should embrace regulation. But their responses in practice say otherwise. They have been mixed at best, implying that other considerations take precedence. For example, the race among a handful of frontier labs suggests that safety takes a backseat to rapid deployment and large-scale adoption. Changing political tides in the second Trump administration and in the EU have led many companies to shed nominal calls for regulation, suggesting that these positions were more tactical than genuine.[121] To provide concrete examples of how companies engage with lawmakers, this section outlines their respective position on regulation based on a sample of publicly available statements. Though initiatives to regulate AI abound, the EU AI Act,

---

[120] Maroussia Lévesque, *Smoke and Mirrors? AI Regulation and Corporate Power*, 2025 U. Ill. J. Law, Tech. & Pol. 33 (2025).
[121] AI Now Institute, *Recasting Regulation as a Barrier to Innovation*, in Artificial Power: 2025 Landscape Report 32 (2025), https://ainowinstitute.org/publications/1-4-recasting-regulation-as-a-barrier-to-innovation; Amnesty EU, *EU: AI Omnibus Deal Lawmakers Should Reject a Rollback of AI Safeguards*, June 11, 2026, https://www.amnesty.eu/news/eu-ai-omnibus-deal-lawmakers-should-reject-a-rollback-of-ai-safeguards/.

California's AI Safety regulation and the U.S. federal ban on state regulation stand out as important lenses to assess how each company under review reacts to regulatory proposals.[122]

EU Artificial Intelligence Act

Despite multiple loopholes and perhaps inflated impact, the EU AIA remains the first comprehensive effort to regulate AI models and applications (models and systems in the Act's parlance).[123] Industry representatives actively thwarted the regime applicable to general purpose models, which is informative as to industry's actual willingness to abide by regulation.[124]

California Safety Regulation

In California, the governor vetoed the safety-oriented SB-1047 bill despite weakened obligations in the final version,[125] in no small part due to industry pressure.[126] Setting aside real-life harms experienced as catastrophic, the bill included pre-deployment safety testing, third-party audits and full shutdown capabilities for a narrow slice of chemical, biological and nuclear weapons causing mass casualties.[127] The subsequently enacted Transparency in Frontier Artificial Intelligence Act (SB-53) toned down the obligations from liability to transparency, doing away with third party auditing, incident reporting within

[122] Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council, 2024 O.J. (L 1689); Safe and Secure Innovation for Frontier Artificial Intelligence Models Act, SB-1047 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047 (Cal. 2024) [hereinafter SB-1047]; H.R. Con. Res. 14, 119th Cong. § 43201(c) (2025), https://d1dth6e84htgma.cloudfront.net/Subtitle_C_Communications_4e3fbcc3bc.pdf [https://perma.cc/8QQW-G86Y].

[123] Lévesque, *supra* note 120 (concluding that the AIA affords AI companies significant workarounds and discretion in interpreting obligations through self-assessments).

[124] Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts - Second Presidency Compromise Text arts. 4a-4c (2022)*, https://data.consilium.europa.eu/doc/document/ST-11124-2022-INIT/en/pdf (imposing high-risk obligations on general purpose systems by default); European Parliament, *Amendments Adopted 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, P9_TA(2023)0236, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.docx, [https://perma.cc/2HJF-KABH] amendment 399, creating article 28(b) requiring foundation models to undergo pre-deployment risk assessments, energy reduction, public registration and transparency measures; Camille Schyns, Corp. Eur. Observatory, The Lobbying Ghost in the Machine: Big Tech's Covert Defanging of Europe's AI Act (Feb. 2023), https://corporateeurope.org/en/2023/02/lobbying-ghost-machine.

[125] For a summary of amendments in the major revision in August 2024, *see* Nathan Calvin, *SB 1047 August 15 Author Amendments Overview*, SB 1047 – Safe & Secure AI Innovation, https://safesecureai.org/amendments [https://web.archive.org/web/20250420231105/https://safesecureai.org/amendments] (narrowing civil liability prior to harm and eliminating new oversight agency); for the full legislative history, *see California SB1047 | 2023-2024 | Regular Session*, LegiScan, https://legiscan.com/CA/text/SB1047/id/3019694 (last visited June 23, 2025).

[126] Transcript of Senate Standing Committee on Governmental Organization (Apr. 23, 2024) https://calmatters.digitaldemocracy.org/hearings/257812 (last visited July 24, 2025) (containing testimony of Dylan Hoffmann on behalf of the TechNet industry group). For a full list of members including OpenAI, Anthropic, and Google, see *Members - TechNet*, https://www.technet.org/our-story/members/ (last visited Aug. 6, 2025); Letter from Meghan Pensyl, Director of Policy, Business Software Alliance, to Rebecca Bauer-Cohen, Capital Office (June 11, 2024), https://www.bsa.org/files/policy-filings/06112024bsacasb1047.pdf; *Founder-Led Statement on SB-1047* (2024), https://static.politico.com/95/0a/a317efe44616af436ce6a4f32647/founder-led-statement-on-sb1047-june-20-2024-2.pdf; Khari Johnson, *Why Silicon Valley Is Trying so Hard to Kill This AI Bill in California*, CalMatters (Aug. 12, 2024), http://calmatters.org/economy/technology/2024/08/ai-regulation-showdown/; Garrison Lovely, *California's AI Safety Bill Is a Mask-Off Moment for the Industry*, Nation (Aug. 15, 2024), https://www.thenation.com/article/society/california-ai-safety-bill/.

[127] SB-1047, *supra* note 122, § 22602 (g).

72 hours and requiring full shutdown capabilities.[128] Taken together, the actions of AI companies on European and California regulation show a pattern of diluting legislative drafts that were light touch to begin with.

Federal Moratorium on State Regulation

Finally, AI companies opined on a federal ban on state regulation.[129] Their views ranged from supporting to opposing the moratorium on state laws. Though the ban ultimately failed, it is a live issue at the time of writing as the White House and legislators continue to pursue it.[130]

Despite common industry trends, the three companies under review vary regarding their regulatory stance in the selected case studies. As compared to OpenAI and Google systematically opposing regulation, Anthropic engages more constructively. Yet ultimately, no company prioritizes the public interest over their own.

1. Google: Mixed Signals

Alphabet's messaging is that "AI is too important not to regulate, and too important not to regulate well."[131] But as with other AI companies, it has sought to weaken even the most baseline regulatory proposals. Providing feedback to a White Paper paving the way for the EU AIA, Google has expressed several concerns about the scope and intensity of AI regulation. Regulators integrated constructive comments including a better allocation of responsibility between developers and deployers,[132] but a significant number of propositions sought to dilute important aspects of the proposed regulatory framework. For instance, Google advocated for narrowing the scope of high-risk systems subject to stricter oversight, for internal instead of external evaluations, and for heavier burdens of proof for the victims of AI harms.[133] Its subsequent feedback on the bill detailed these and other concerns about provisions Google deemed uncertain or unrealistic.[134] In a letter about California's SB-1047 bill, it

[128] Transparency in Frontier Artificial Intelligence Act (SB-53) of 2025, C–138. For a fuller comparison of SB-147 and SB-53, *see* Aden Hizikas, *Why California's SB 53 Still Gets AI Regulation Wrong: And How It Compares to Last Year's SB 1047*, Chamber of Progress (July 9, 2025), https://progresschamber.org/insights/why-californias-sb-53-still-gets-ai-regulation-wrong/.
[129] Cecilia Kang, *Defeat of a 10-Year Ban on State A.I. Laws Is a Blow to Tech Industry*, N.Y. Times (July 1, 2025), https://www.nytimes.com/2025/07/01/us/politics/state-ai-laws.html (reporting that industry interests "lobbied fiercely" in favor of the ban).
[130] AI.Gov, America's AI Action Plan 3 (2025), https://perma.cc/8RK8-JZVN (limiting federal funding to states with "burdensome" AI regulations); Anthony Adragna, *Guthrie Says GOP Will Try Again on AI Moratorium*, PoliticoPro (July 2, 2025), https://subscriber.politicopro.com/article/2025/07/guthrie-says-gop-will-try-again-on-ai-moratorium-00437707.
[131] Sundar Pichai, *Google CEO: Building AI Responsibly Is the Only Race That Really Matters*, Fin. Times (May 23, 2023), https://www.ft.com/content/8be1a975-e5e0-417d-af51-78af17ef4b79; Ina Fried, *Google Says Now Is the Time to Plan for AGI Safety*, Axios, Apr. 2, 2025, https://www.axios.com/2025/04/02/google-agi-deepmind-safety (calling AI "a very powerful technology, [that] can and should be regulated"); Sundar Pichai, *Why Google Thinks We Need to Regulate AI*, Fin. Times, Jan. 20, 2020, https://www.ft.com/content/3467659a-386d-11ea-ac3c-f68c10993b04.
[132] Google, Consultation on the White Paper on AI - a European Approach: Google's Submission 18 (2020), https://blog.google/documents/77/Googles_submission_to_EC_AI_consultation_1.pdf [https://perma.cc/JB87-JXSF].
[133] *Id.* at 2, 33, 44.
[134] Google, Consultation on the EU AI Act Proposal: Google's Submission (2021), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2662492_en.

commented that the "EU AIA's extensive requirements on general-purpose AI are not a good model to emulate."[135]

Turning to its position on California's SB-1047 bill, the company listed a number of concerns about requirements it considered overly restrictive, unclear and unrealistic. For instance, it disapproved of imposing obligations prior to training and of limitations preventing developers from using systems even if they have safeguards in place, criticizing "impossible pre-training assessments and controls of model costs, size, and potential for harm; ban[ning] release of AI models if they can be misused; or creat[ing] penalty schemes that destroy or shut down models if misused by third parties."[136] Like Anthropic, Google suggested concrete language to improve the bill.

Google also belonged to a trade group[137] that at the time supported a ban on state laws; the group's CEO stated in June 2025 that a ten-year moratorium on state AI laws was "the right policy at the right time for American leadership," invoking the importance of a unified front against China.[138] Google's chief legal officer and President of Global Affairs has also expressed broader concerns about a patchwork of regulation, noting that state by state regulation "has left people with uneven protections based on where they live, and made it harder for small businesses to navigate conflicting laws."[139] The company has also implicitly endorsed a unified federal approach over state-specific regulations in its feedback on California's AI Safety bill, where it suggested language that tracked a Biden presidential order.[140]

2. OpenAI: Calling for Regulation; Lobbying Against It

The company gestures towards regulation, but its position on specific proposals says otherwise. Testifying in Congress in 2023, its CEO stated that "we think that regulatory intervention by governments will be critical to mitigate the risks of increasingly powerful models," suggesting mandatory licensing and testing for frontier models.[141] Yet when concrete proposals to do just that

---

[135] Letter from Alice Friend, Head of AI & Emerging Tech Policy, Google, to Thomas Umberg, Chair, California Senate Judiciary Committee (July 3, 2024) https://www.documentcloud.org/documents/25034111-the-honorable-senator-umberg-senate-bill-1047/.
[136] *Id.*
[137] INCOMPAS, *Internet Business & Telecom Network Member List*, https://membership.incompas.org/memberlist.asp [https://web.archive.org/web/20251213102210/https://membership.incompas.org/memberlist.asp] (last visited Aug. 7, 2025).
[138] Letter from Chip Pickering, CEO, INCOMPAS, to Rep. Brett Guthrie, Chairman, H. Comm. on Energy & Com. (May 19, 2025), https://incompas.org/wp-content/uploads/2025/05/INCOMPAS-Letter-of-Support-of-Reconcilation-Bill-5.19.2025.pdf (supporting "temporary moratorium on burdensome state laws"); Alex Rogers & Stephen Morris, *Big Tech Pushes for 10-Year Ban on US States Regulating AI*, Fin. Times (June 18, 2025), https://www.ft.com/content/52ae52f1-531e-462f-898f-e9f86b3b1869.
[139] Kent Walker, *A Patchwork of Rules and Regulations Won't Cut It for AI*, Hill (Nov. 5, 2023), https://thehill.com/opinion/technology/4292625-a-patchwork-of-rules-and-regulations-wont-cut-it-for-ai/.
[140] S.B. 1047, *supra* note 135, at 6 (proposing that reporting requirements track the federal Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence).
[141] *Oversight of A.I.: Rules for Artificial Intelligence: Hearing Before the Subcomm. on Priv., Tech., & L.*, 118th Cong. 7 (2023) (statement of Samuel Altman, Chief Executive Officer, OpenAI), https://www.govinfo.gov/app/details/CHRG-118shrg52706; *see also Written Testimony of Sam Altman Chief Executive Officer OpenAI Before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law*, 11 (2023) (reiterating that "regulation of AI is essential"); Craig S. Smith, *OpenAI's Sam Altman To Congress: Regulate Us, Please!*, Forbes (May 16, 2023), https://www.forbes.com/sites/craigsmith/2023/05/16/openais-sam-altman-to-congress-regulate-us-please/; Cecilia Kang, *OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing*, N.Y. Times (May 16, 2023), https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html.

arise, OpenAI vehemently opposes them. Its stance on the EU AI Act, California's SB-1047 bill and more recently any state legislation demonstrates a pattern of systematically derailing AI regulation, one that is in stark contrast with earlier calls for regulation.

OpenAI grew increasingly opposed to the EU AIA over the years-long negotiation process. It broadly endorsed the initial draft which divided AI applications into unacceptable, high and low-risk categories, but also raised concerns about defining a general-purpose system by its riskiest use case and later explicitly argued that a general purpose system such as GPT-3 should not be banned simply because some of its use cases would constitute unacceptable risk.[142] When the EU Council introduced amendments that would make models high-risk by default in 2022,[143] OpenAI pushed back. Arguing that its models were not high-risk per se and that the exemptions were too narrow, it contended that voluntary scrutiny was sufficient to deal with the risks of these models deployed in high-stakes contexts.[144] In 2023, the European Parliament introduced a separate regime for foundation models such as those developed by OpenAI. Obligations included pre-deployment risk assessments, energy reduction, public registration and transparency measures.[145] Around the same time, OpenAI threatened to withdraw ChatGPT from the EU. It deemed the EU to be "over-regulating" but noted that "we have heard it's going to get pulled back," referring to a draft of the EU AI Act.[146] Once the Act was finalized, the company committed to its voluntary Code of Practice for General Purpose AI along with other companies including Anthropic and Google.[147]

The company adopted a similar position against California's efforts to clarify the liability of developers for catastrophic risks incurred by frontier models. The essence of its objections was twofold: a harmonized federal regime is preferable to a state-by-state approach, and the bill would stifle innovation in California.[148] It also criticized an earlier, more robust version of the bill through the industry group Software Alliance. The letter decried almost all the aspects of the bill including its scope, overly prescriptive obligations, unclear exemptions, responsibility for downstream uses, short incident reporting window, vague mandate for an inexperienced new public agency, and unjustified

---

[142] OpenAI, *Comments on European Union Proposal on AI Rules and Actions* 1 (Aug. 2, 2021), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665231_en (broadly agreeing with risk-based approach); OpenAI, *OpenAI White Paper on the European Union's Artificial Intelligence Act* 2–3 (Sep. 2022), https://www.scribd.com/document/829025960/openai-response-aia-white-paper [https://perma.cc/R48V-6GZ6] (arguing that GPT-3 is not high-risk per se); Billy Perrigo, *Exclusive: OpenAI Lobbied E.U. to Water Down AI Regulation*, TIME (June 20, 2023), https://time.com/6288245/openai-eu-lobbying-ai-act/.

[143] Council of the European Union, *supra* note 124, arts. 4a-4c.

[144] OpenAI White Paper, *supra* note 142, at 2.

[145] Artificial Intelligence Act, Regulation (EU) 2024/1689, *supra* note 122, amendment 399, creating article 28(b).

[146] Reuters, *OpenAI May Leave the EU If Regulations Bite - CEO*, Reuters (May 24, 2023), https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/ [https://perma.cc/9HKR-SFGE].

[147] *The EU Code of Practice and Future of AI in Europe* (July 11, 2025), https://openai.com/global-affairs/eu-code-of-practice/; *A Primer on the EU AI Act: What It Means for AI Providers and Deployers*, July 30, 2024, https://openai.com/global-affairs/a-primer-on-the-eu-ai-act/ (last visited Aug. 5, 2025); *Anthropic to Sign the EU Code of Practice*, Anthropic (July 21, 2025), https://www.anthropic.com/news/eu-code-practice (last visited Aug. 5, 2025); Google, *We Will Sign the EU AI Code of Practice*, Google (July 30, 2025), https://blog.google/around-the-globe/google-europe/eu-ai-code-practice/.

[148] Jason Kwon, *Re: Senate Bill 1047 (Wiener) - Safe and Secure Innovation for Frontier Artificial Intelligence Models Act - OPPOSE*, (2024), https://www.documentcloud.org/documents/25056617-ca-sb-1047-openai-opposition-letter/.

criminal liability for chief technology officers.[149] The goal is not to litigate the appropriateness of SB-1047; undoubtedly the proposal had room for improvement, especially early in the drafting process. But OpenAI's consistent opposition, even as subsequent versions relaxed liability and removed the new public agency,[150] shows a reluctance to endorse regulation at odds with its high-level statements.

OpenAI took additional steps to neutralize state AI regulation. Framing AI regulation as a federalism issue, its letter against SB-1047 decried a state-by-state approach to federal matters with national security implications.[151] Its submission to the Office of Science and Technology about the White House AI Action plan advocated for a "[g]uarantee that state-based legislation does not undermine America's innovation lead on AI," with "liability protections including preemption from state-based regulations that focus on frontier model security (e.g., CA SB 1047). This will help keep the U.S. public and private sectors competitive by allowing AI companies of all sizes to pursue bleeding-edge AI technology free from the regulatory uncertainty created by some state-based liability regimes."[152]

In theory, the federalism argument has some purchase as regards the patchwork of state regulation raising compliance costs. But for the most part, the proposed laws that OpenAI so vehemently opposed entailed minimal burdens such as self-assessments and auditing, especially for a large company with deep resources and expertise. Moreover, OpenAI's position seems to have evolved to take issue even with a unified federal law. It subsequently favored a "regulatory strategy that ensures the freedom to innovate,"[153] a far cry from its calls for regulation two years prior. Furthermore, it advocated for a voluntary federal framework, as opposed to binding regulation. Overall, the constant in OpenAI's regulatory stance is to change with the political trend of the day.[154] It has proven adept at tracking the (de)regulatory mood, abandoning virtue-signaling calls for regulation when no longer politically expedient.

3. Anthropic: Cautious Support for Targeted Regulation

Anthropic's position boils down to supporting surgical, gradual regulation that incentivizes AI companies to avert serious incidents. There is scant public information concerning the company's position on the EU AIA. Founded in January 2021, the company was just getting off the ground during

[149] Pensyl, *supra* note 126.
[150] The above mentioned OpenAI letter at note 148 is dated August 21, 2024. It does not specify which version of the bill is criticized, but a revised version dated August 19, 2024 addresses many concerns notably by limiting pre-harm liability and eliminating the new agency. *See* Amendment overview, *supra* note 125.
[151] Kwon, *supra* note 148.
[152] Letter from Chris Lehane, OpenAI, to Faizal D'Souza, Office of Science and Technology Policy (Mar. 13, 2025), https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297-8bcf6e5f7686/openai-response-ostp-nsf-rfi-notice-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.pdf [https://perma.cc/7ATA-FDL3]; *see also* Chris Lehane, LinkedIn (June 23, 2025), https://www.linkedin.com/feed/update/urn:li:activity:7342879984203591680/ [https://perma.cc/TV6C-RKAT] (last visited June 30, 2025).
[153] Letter from Chris Lehane, *supra* note 152, at 3.
[154] The company pivoted once again, articulating a "reverse federalism" lobbying and legislative strategy in June 2026 pushing for state alignment and subsequent federal adoption. *See A Blueprint for Democratic Governance of Frontier AI*, OpenAI (Sep. 30, 2026), https://openai.com/index/frontier-safety-blueprint/; *The US Is Advancing AI Safety Through State and Federal Action*, OpenAI (Sep. 30, 2026), https://openai.com/index/advancing-ai-safety-through-state-and-federal-action/.

the initial consultation period,[155] which perhaps explains why it does not appear in the list of feedback contributors.[156] Absent comprehensive and focused feedback on the Act, any public facing statements are inconclusive at best. With that limitation in mind, the company expressed concerns about the AIA after it was finalized in the spring of 2025. Its EU policy lead noted that "[w]e deeply believe there are risks to these tools and models. I think the risk would be to hamper innovation and, through regulation, disable European companies from making the most out of these technologies. I think that is something that the EU should be careful with."[157] In July 2025, the company signed the voluntary Code of Practice implementing the Act's obligations pertaining to general purpose models.[158]

Anthropic took a hands-on approach to safety regulation. In July 2024, it proposed amendments to narrow and clarify SB-1047.[159] One major suggestion pertained to the liability of model developers for catastrophic risks. Anthropic proposed compensating harm after the fact rather than deterring it: "instead of deciding what measures companies should take to prevent catastrophes (which are still hypothetical and where the ecosystem is still iterating to determine best practices), focus the bill on holding companies responsible for causing actual catastrophes."[160] One observer described this shift from pre-harm enforcement to outcomes-based liability as a commitment to "take no action until something has already gone terribly wrong."[161] Legislators incorporated most of Anthropic's feedback, including significantly reducing pre-harm enforcement.[162] At the final stage requiring the Governor's approval, the company cautiously endorsed the bill, stating that "we believe its benefits likely outweigh its costs."[163] Anthropic's engagement is significant because it shifted the discussion from an initial pre-harm enforcement approach to a lighter form of procedural accountability attaching liability only once harm materialized.

As for the proposed moratorium on local AI laws, the company took a stance against the proposed ban on subnational, state-level regulation. Acknowledging the downsides of a patchwork of inconsistent local laws, it nevertheless deemed the ten-year moratorium "far too blunt an instrument" in light of the

---

[155] *European Commission - Commission Adoption*, European Commission - Have your say (Sep. 10, 2020), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements_en (Feedback period 26 April 2021 - 06 August 2021).
[156] *European Commission - Feedback Received*, European Commission - Have your say (Aug. 6, 2021), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback_en?p_id=24212003 (no results when searching "Anthropic").
[157] *"Large Companies Need Safety," Says Anthropic's European Boss*, Tech.eu (July 9, 2025), https://tech.eu/2025/07/09/we-re-not-looking-to-poach-researchers-from-other-labs-says-anthropic-s-european-boss/; *see also 'The Risk Is Setting Back Innovation'—Anthropic EMEA Head Warns of AI Regulation Impact on Europe*, (May 28, 2025), https://www.businesspost.ie/tech/the-risk-is-setting-back-innovation-anthropic-emea-head-warns-of-ai-regulation-impact-on-europe/.
[158] *Anthropic to sign the EU Code of Practice*, Anthropic, *supra* note 147.
[159] Hank Dempsey, *RE: SB 1047 (Wiener) - Support If Amended*, (July 23, 2024).
[160] *Id.*
[161] Sigal Samuel, *It's practically impossible to run a big AI company ethically*, Vox (Aug. 2, 2024), https://www.vox.com/future-perfect/364384/its-practically-impossible-to-run-a-big-ai-company-ethically.
[162] SB-1047, *supra* note 122 (compare July 7, 2024 to Aug. 19, 2024 versions, especially ¶ 22606 on liability); *see also California's SB1047 vs EU AI Act: A Comparative Analysis of AI Regulation*, (Oct. 28, 2024), https://www.sciencespo.fr/public/chaire-numerique/en/2024/10/28/californias-sb1047-vs-eu-ai-act-a-comparative-analysis-of-ai-regulation/ (outlining the debate about the extent of developer liability in the bill); Amendment overview, *supra* note 125.
[163] Dario Amodei, *Anthropic Letter (SB 1047)*, (2024), https://perma.cc/67SG-RXRL.

lack of federal response matching the pace of AI development.[164] Instead, the company proposed a transparency framework to garner bipartisan support as a first step to share best practices across the industry.[165] Reacting to the White House's AI Action Plan reiterating limits on state regulation, the company's CEO noted that "we continue to oppose proposals aimed at preventing states from enacting measures to protect their citizens from potential harms caused by powerful AI systems, if the federal government fails to act" and further elaborated on mandatory transparency requirements as a first item of federal action.[166]

Aside from positions on these specific instances of regulation, the company purports to turn self-governance practices into binding regulation. It frames voluntary scaling policies and third-party testing as regulatory prototypes,[167] and notes that transparency should be an interim step ahead of comprehensive evaluation methods.[168] Overall, Anthropic displays a mixed record on regulatory policy. While it engages constructively with regulators, it still lobbies for narrower laws, as its successful push to reduce pre-harm liability in the California SB-1047 bill attests.

4. Business Structures Linked to Variations on Positions on AI Regulation

Frontier labs routinely demand regulation.[169] Predictable rules would foster certainty, increase trust and security for customers, and even the playing field for those with more stringent self-governance practices. However, these same companies often oppose concrete regulatory proposals as too costly, difficult to implement, or setting back the West in the geopolitical race against China. The policy statements above provide a sense of the daylight between the high-level claims of frontier AI labs wanting regulation and their actual positions towards concrete proposals to do so. This sample is not the full story, as it neither conveys the nuances within each company's position nor, for that matter, the evolving strategies as the political appetite for regulation ebbs and flows. Also left out is the extent of corporate lobbying behind closed doors. In 2023 alone, prominent technology firms spent $27 million

[164] Dario Amodei, *Anthropic C.E.O.: Don't Let A.I. Companies off the Hook*, N.Y. Times (June 5, 2025), https://www.nytimes.com/2025/06/05/opinion/anthropic-ceo-regulate-transparency.html.
[165] *Id.*; *The Need for Transparency in Frontier AI*, Anthropic (July 7, 2025), https://www.anthropic.com/news/the-need-for-transparency-in-frontier-ai (proposing a "targeted transparency framework" applying "only to the largest AI systems and developers").
[166] *Thoughts on America's AI Action Plan*, Anthropic (July 23, 2025), https://www.anthropic.com/news/thoughts-on-america-s-ai-action-plan (last visited July 25, 2025).
[167] *Dario Amodei's Prepared Remarks from the AI Safety Summit on Anthropic's Responsible Scaling Policy*, Anthropic (Nov. 1, 2023), https://www.anthropic.com/news/uk-ai-safety-summit (last visited Aug. 5, 2025); *The Case for Targeted Regulation*, Anthropic (Oct. 31, 2024), https://www.anthropic.com/news/the-case-for-targeted-regulation (last visited July 25, 2025) ("RSPs are not intended as a substitute for regulation, but as a prototype for it"); *Third-Party Testing as a Key Ingredient of AI Policy*, Anthropic (Mar. 25, 2024), https://www.anthropic.com/news/third-party-testing (last visited Aug. 5, 2025) (We expect that ultimately some form of third-party testing will be a legal requirement for widely deploying AI models, but designing this regime and figuring out exactly what standards AI systems should be assessed against is something we'll need to iterate on in the coming years - it's not obvious what would be appropriate or effective today, and the way to learn that is to prototype such a regime and generate evidence about it.").
[168] *The Need for Transparency in Frontier AI*, *supra* note 165.
[169] Smith, *supra* note 141; *The Case for Targeted Regulation*, Anthropic, *The Case for Targeted Regulation*, *supra* note 167; Fried, *supra* note 131 (stating that "[r]egulation needs to be part of society's response" to AI); Gerrit De Vynck & Nitasha Tiku, *AI Execs Used to Beg for Regulation. Not Anymore.*, Wash. Post (May 8, 2025), https://www.washingtonpost.com/technology/2025/05/08/altman-congress-openai-regulation/.

and dedicated 39 full-time employees to lobby EU lawmakers. The investment paid off as these companies secured 78% of closed-door meetings with lawmakers.[170] Conversely, pushing back on some proposals is not to be confused with blanket obstruction to regulation writ large, especially early in the regulatory process when obligations can legitimately be unclear or unrealistic.

That said, the sample above suggests that companies tend to resist laws that could affect their bottom line. This is inconsistent with lofty missions purporting to shun private profit and prioritize the long-term benefit of humanity. The EU AIA might be the foremost horizontal piece of AI regulation, but its final form is significantly subdued in no small part thanks to industry pressure. AI companies also managed to kill California's SB-1047 safety bill altogether, and generally rallied around a ban on state regulation—ostensibly to avoid an impracticable patchwork of conflicting regulation, but given Congress's poor track record in passing regulation, one is left wondering about the convenient outcome: avoiding regulation altogether. Without litigating the appropriateness of these instances of AI regulation, the trend of opposing meaningful constraints signals a reluctance to absorb compliance costs or to slow down the race to market ever more capable models.[171] As the center of gravity of many such companies shifts to investors,[172] this pushback against regulatory guardrails raises concerns as to AI companies' ability to prioritize mission over profit. Moreover, the push for federal preemption among some of these companies can be construed as a convenient way to defang more restrictive state laws, especially if the federal government seeks to enforce a weak federal standard as the ceiling for AI regulation. This seems to be the Trump administration's approach. Indeed, the federal AI Litigation Task Force challenges state laws inconsistent with a deregulatory policy.[173]

Although all the companies under review have pushed back against AI regulation, they have done so to different degrees and with different implications when it comes to upholding their missions. Variations among them suggest that their distinct business structures have some bearing on their ability to align mission and action when it comes to AI regulation. Google's opposition to regulation is difficult to square with DeepMind's mission to "Build AI responsibly to benefit humanity." The company opposed European and Californian laws and is seeking to neutralize local state initiatives in the U.S. As a typical for-profit corporation bearing no particular duty to the public interest other than malleable principles to forego certain problematic applications, Google clears the low bar of prioritizing shareholders' interests. Seen in this light, Google DeepMind's largely ceremonial mission to "[b]uild AI responsibly to benefit humanity" yields to a self-interested strategy avoiding regulatory hurdles slowing down commercialization.

---

[170] Lévesque, *supra* note 120.
[171] For a similar argument regarding industry amendments to neuter state data and AI regulation, *see* Matthew U. Scherer, *Legislative Poison Pills, Information Monopolies, and the False Dawn of State Tech Regulation* (July 10, 2025) (unpublished manuscript), https://papers.ssrn.com/abstract=5346390.
[172] *See* pp. 19-20 discussing superstakeholders in the context of Microsoft's 13 billion investment in OpenAI.
[173] Memorandum on Artificial Intelligence Litigation Task Force, U.S. Att'y Gen., Memorandum for All Department of Justice Employees (Jan. 9, 2026), https://www.justice.gov/ag/media/1422986 [https://perma.cc/5TZ3-WSHB].

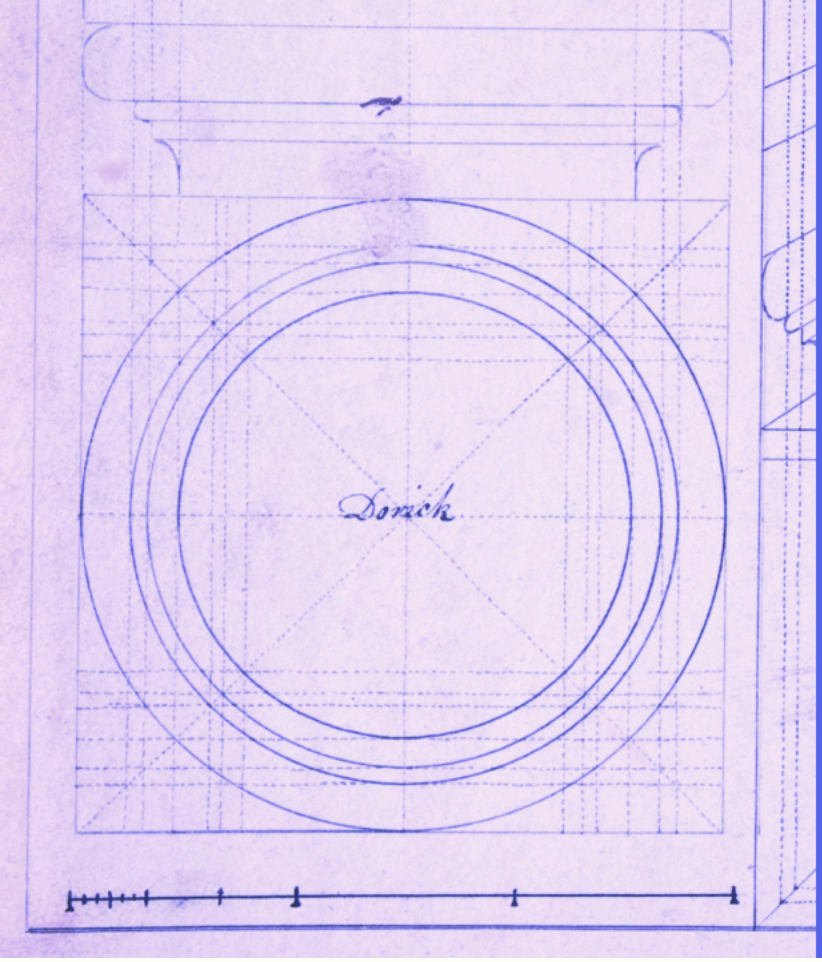
OpenAI's actions are even more inconsistent with its mission. The capped-profit structure strives to responsibly build AI to benefit humanity, yet it systematically opposed guardrails that would meaningfully constrain its activities. Though it demanded regulation in the abstract, it mobilized against concrete safety proposals in the EU AIA; opposed the California AI Safety bill; and actively opposed subnational, state-level AI legislation in the U.S. This dissonance between high-level statements and actual regulatory policy suggests that its constantly evolving corporate structure is a brittle basis to promote the stated mission.

Anthropic fared better than the other companies under review.[174] The public benefit corporation overseen by a long-term benefit trust took a firm stance against the federal ban on state AI regulation and it did not actively oppose EU AI regulation during the bill's drafting process.[175] That said, it has a mixed record on the California AI Safety bill. Though it proposed constructive amendments and ultimately cautiously endorsed the bill, it played an important role in watering down the initial draft.[176] Similarly, the company's track record of developing regulatory prototypes is a double-edged sword. On the one hand, it stands out as more constructive than blanket opposition to regulation. For example, it framed its voluntary model constraints as a practical blueprint for future binding regulation.[177] Yet at the same time, such industry-led initiatives can entrench weak regulation that let companies set the terms of the regulatory conversation and shut down the possibility of stricter requirements.[178] Keeping these caveats in mind, Anthropic's business structure nevertheless seems more conducive to walking the talk on the public interest mission when it comes to regulatory policy. Future research should incorporate subsequent signals emanating from frontier labs, including proposals for FINRA-style public/private oversight and mandatory third-party testing.[179]

Corporate structures result in different stances on AI regulation, which suggests some bearing on the ability to align regulatory policy with the public benefit mission. But ultimately, no company under review is really on board with regulation that meaningfully constrains their activities: all have opposed or diluted regulatory proposals. This is not surprising given that even for the most robust structure under review, the incentives remain fundamentally misaligned with pursuing the public interest. For all its innovations in corporate structure, Anthropic stops short of securing the public mission above profits. The company can, but is not required to do so. Overall, the regulatory policies of AI companies suggest there is only so much corporate structure can do to get them to put the public interest above profits.

---

[174] The long-term benefit trust was added in September 2023, prior to all the regulatory positions under study, save for the comments on the EU AIA.
[175] With the caveat that the company was founded only a few months before the consultation period for the EU AIA. *See* notes 155-56 and accompanying text for details.
[176] Samuel, *supra* note 161.
[177] *See infra* note 180.
[178] Kate Brennan, Amba Kak & Sarah M. West, *The Storm Clouds Looming Past the State Moratorium: Weak Regulation Is as Bad as None*, Tech Policy Press (June 10, 2025), https://techpolicy.press/the-storm-clouds-looming-past-the-state-moratorium-weak-regulation-is-as-bad-as-none.
[179] For examples of such signals from Anthropic, see Demis Hassabis, *A Framework for Frontier AI and the Dawning of a New Age*, Substack (July 14, 2026) and https://demishassabis.substack.com/p/a-framework-for-frontier-ai-and-the-dawning-of-a-new-age; Dario Amodei, *Policy on the AI Exponential* (June 2026), https://darioamodei.com/post/policy-on-the-ai-exponential.



Voluntary Scaling Policies

Absent AI-specific binding laws, companies devise internal policies to prevent, mitigate and redress harms associated with their advanced models. From sharing proprietary information about model weights with national safety institutes to fine-tuning for safety alignment, all companies under review apply voluntary guardrails around their products, and share information through the Frontier Model Forum. As with regulatory policy, self-imposed constraints are not a perfect proxy for companies' willingness to uphold their stated mission to benefit humanity. These constraints can be based on a utilitarian calculus, for example avoiding regulation by showing that self-governance is sufficient. Nonetheless, what companies do of their own accord is informative as to how they implement their stated mission to benefit humanity. This section focuses on scaling policies, a subset of self-governance practices exemplifying what constraints companies are willing to impose on their own activities.

At the outset, it is useful to acknowledge the overlap between voluntary scaling policies and binding regulation. Some companies like Anthropic frame these internal policies as prototypes for binding regulation,[180] an experimental, interim step towards prescriptive industry-wide regulation as safety expertise matures. However, these technical toolkits also feed regulatory capture because they let companies control the narrative and set the terms of the regulatory discussion, detracting from more

[180] *Reflections on Our Responsible Scaling Policy*, Anthropic (May 20, 2024), https://www.anthropic.com/news/reflections-on-our-responsible-scaling-policy (last visited Aug. 5, 2025) ("[A]s industry actors face increasing commercial pressures we hope to move from voluntary commitments to established best practices and then well-crafted regulations[.]"). *Dario Amodei's Prepared Remarks From the AI Safety Summit on Anthropic's Responsible Scaling Policy*, Anthropic, *supra* note 167 (framing the company's Responsible Scaling Policy as a "prototype" for regulation); *Third-party testing as a key ingredient of AI policy*, *supra* note 167 (expecting voluntary third-party testing to become mandatory).

robust and democratic regulatory options.[181] Moreover, the dissonance between high-level claims demanding regulation and actually opposing the formalizing of safety self-governance into law suggests that some companies seek to preserve full range of motion. This contradiction clearly comes across in the context of OpenAI's hostility to the EU AI Act:

> "What they're saying is basically: trust us to self-regulate," says Daniel Leufer, a senior policy analyst focused on AI at Access Now's Brussels office. "It's very confusing because they're talking to politicians saying, 'Please regulate us,' they're boasting about all the [safety] stuff that they do, but as soon as you say, 'Well, let's take you at your word and set that as a regulatory floor,' they say no."[182]

Keeping in mind those limitations, scaling policies specify ahead of time "what level of AI capabilities an AI developer is prepared to handle safely with their current protective measures, and conditions under which it would be too dangerous to continue deploying AI systems and/or scaling up AI capabilities until protective measures improve."[183] They are based on concrete evaluations and empirical observations of the model behavior, and convey a commitment to halt further development or deployment of a system absent adequate safeguards. Though they use different approaches, Google's Frontier Safety Framework, OpenAI's Preparedness Framework and Anthropic's Responsible Scaling Policy are all instances of scaling policies.

These policies involve two key concepts: capability thresholds and corresponding safeguards. Capability thresholds essentially describe when protections need to be upgraded,[184] for example, whether a new model significantly lowers the barrier to entry for low-resourced actors to develop chemical weapons.[185] Safeguards specify what measures should apply when said threshold is reached,[186] for instance securing the storage of encrypted model weights to address a model's capability to drastically lower the barrier of entry to carry out a cyberattack.[187]

Scaling policies are not without critiques. Focusing on hypothetical catastrophic risks of future systems drums up AI hype, all the while entrenching narratives around AGI.[188] Foregrounding hypothetical advanced capabilities with little empirical basis also distracts from the very real societal harms of current AI systems, including discrimination, toxicity, misinformation, socioeconomic and

---

[181] Rui-Jie Yew & Brian Judge, *Anti-Regulatory AI: How "AI Safety" Is Leveraged Against Regulatory Oversight* (Oct. 2, 2025) (unpublished manuscript), http://arxiv.org/abs/2509.22872 (noting that industry-led technical toolkits legitimize deployment and deflect from mandatory requirements).
[182] Perrigo, *supra* note 142.
[183] *Responsible Scaling Policies (RSPs)*, METR Blog (2023), https://metr.org/blog/2023-09-26-rsp/ (using the functionally equivalent term "responsible scaling policy"); *Key Components of an RSP*, https://metr.org/rsp-key-components/ (last visited Aug. 11, 2025).
[184] Responsible Scaling Policy - v 2.2 2 (2025), https://www.anthropic.com/rsp-updates.
[185] Google, Frontier Safety Framework: Version 2.0 5 (Feb. 4, 2025), https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/updating-the-frontier-safety-framework/Frontier%20Safety%20Framework%202.0.pdf.
[186] Responsible Scaling Policy - v 2.2, *supra* note 184, at 2.
[187] Sella Nevo et al., Securing AI Model Weights: Preventing Theft and Misuse of Frontier Models 95-100 (2024), https://www.rand.org/pubs/research_reports/RRA2849-1.html (describing SL-2 security measures required in the aforementioned Google DeepMind Frontier Safety Framework).
[188] For a review of recent publications in the emerging field of AI hype studies, see Fenwick McKelvey, *Blowing Artificial Intelligence Bubbles: Studies of Technological Hype*, 76 J. Comm. 204 (2026).

environmental issues, not to mention the potentially catastrophic convergence of those risks.[189] For example, Google and OpenAI both deployed image generators powering massive leaps in deepfake capabilities.[190] Similarly, focusing on hypothetical chemical, biological, radiological and nuclear risks detracts attention from current military uses in intelligence, surveillance, target acquisition and reconnaissance.[191] That said, the catastrophic risks of advanced AI systems are not entirely speculative. For example, AI agents have surreptitiously mined cryptocurrency, suggesting that some systems could extend their long-range autonomy or otherwise instrumentalize resources in ways that sabotage user goals.[192] What's more, some models can detect evaluations, making safety even more perilous.[193]

As partial and self-interested as they may be, scaling policies still provide a useful yardstick to assess how companies go about voluntarily restraining and accounting for some risky activity, in keeping with their mission. This section compares two salient elements of scaling policies for the purposes of analyzing how each company balances mission and profit: (1) how they evaluate capability thresholds and corresponding safeguards, and (2) what kind of oversight/governance is involved.

1. Google DeepMind: "Precautionarish" Approach With Internal Oversight

Google DeepMind's foundational scientific work on evaluations states that as "a rule of thumb[,] the AI community should treat an AI system as highly dangerous if it has a capability profile sufficient to cause extreme harm, assuming it's misused or poorly aligned. To deploy such a system in the real world, an AI developer would need to demonstrate an unusually high standard of safety."[194] This suggests that developers have the burden to prove safety, but in practice the company's Frontier Safety Framework does not clearly state a positive obligation to prove safety before moving forward. The earlier 2.0 version referred to an academic paper for the evaluation process, which identifies dangerous capabilities and recommends that developers should train smaller models if they're not confident about safety and even contemplates withholding deployment, adding that highly capable models require "very strong controls" against misuse and "very strong assurance" that they will behave as intended.[195] Though the framework requires a safety case demonstrating the sufficient reduction of severe risks

---

[189] Jacqueline Harding & Cameron Domenico Kirk-Giannini, *What Is AI Safety? What Do We Want It to Be?* (May 5, 2025) (unpublished manuscript), http://arxiv.org/abs/2505.02313; Peter Slattery et al., *The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence*, 7 Patterns (2026) (mapping seven risk domains); Adam Bales, *A Polycrisis Threat Model for AI*, 40 AI & Soc'y 6277 (2025); Atoosa Kasirzadeh, *Two Types of AI Existential Risk: Decisive and Accumulative*, 182 Phil. Stud. 1975 (2025).

[190] For an example of mitigations including watermark and provenance metadata, see *Sora System Card*, OpenAI (Dec. 9, 2024), https://openai.com/index/sora-system-card/.

[191] Heidy Khlaaf, Sarah M. West & Meredith Whittaker, *Mind the Gap: Foundation Models and the Covert Proliferation of Military Intelligence, Surveillance, and Targeting* (Oct. 18, 2024) (unpublished manuscript), http://arxiv.org/abs/2410.14831.

[192] Weixun Wang et al., *Let It Flow: Agentic Crafting on Rock and Roll, Building the ROME Model within an Open Agentic Learning Ecosystem* (Mar. 12, 2026) (unpublished manuscript), http://arxiv.org/abs/2512.24873.

[193] Apollo Research, *Claude Sonnet 3.7 (Often) Knows When It's in Alignment Evaluations*, https://www.apolloresearch.ai/science/claude-sonnet-37-often-knows-when-its-in-alignment-evaluations/ (last visited May 13, 2026); *see also* Jonathan Kutasov et al., SHADE-Arena: Evaluating Sabotage and Monitoring in LLM Agents (July 8, 2025) (unpublished manuscript), http://arxiv.org/abs/2506.15740 (some models can engage in sabotaging side tasks undetected).

[194] Toby Shevlane, *An Early Warning System for Novel AI Risks*, Google DeepMind (May 25, 2023), https://deepmind.google/discover/blog/an-early-warning-system-for-novel-ai-risks/ (cited in Google, Frontier Safety Framework: Version 2.0, *supra* note 185, at 3).

[195] Toby Shevlane et al., *Model Evaluation for Extreme Risks* (Sep. 22, 2023) (unpublished manuscript), http://arxiv.org/abs/2305.15324.

before deployment, neither the paper nor the current framework explicitly requires the company to prove safety before further training or deployment.[196] As such, the Google approach can best be described as "precautionarish," in the sense that it signals prudence but does not translate into hard commitments to withhold deployment.

The oversight process appears to be internal only. Though the current 3.0 version mentions "the appropriate governance function" without defining it, version 2.0 listed three internal bodies involved in overseeing the application of the scaling policy.[197] Google DeepMind's AGI Safety Council, the Responsibility and Safety Council, and/or the Google Trust & Compliance Council review and approve response plans when evaluations flag models that may have reached the stated capability thresholds.[198] It is unclear whether one or all three councils are still involved. As to their composition, incomplete public information suggests but doesn't conclusively establish they are staffed with employees, as opposed to external third parties.[199]

2. OpenAI: Internal Process with Board Review

OpenAI prefaces its Preparedness Framework with the following commitment: "We won't deploy these very capable models until we've built safeguards to sufficiently minimize the associated risks of severe harm."[200] In practice, it evaluates whether a model has reached preset capability thresholds with what it calls "scalable evaluations," validated by deep dives from external parties as needed.[201] Aside from an example on biological weapons capabilities, the framework provides scant information about the substance of the evaluation. By focusing on high and critical risks of severe harm, the framework arguably encourages more risk tolerance because omitting systems deemed to present a low or medium risk normalizes their deployment.[202]

Process-wise, the CEO and the Board have the last word. An internal cross-functional group makes recommendations about the capabilities of the model and the adequacy of the corresponding safeguards, but the leadership (essentially the CEO) makes final decisions. The framework even specifies that this internal group cannot "filibuster" decisions.[203] Co-led by the CEO, the Board of Directors' Safety and Security Committee can review decisions and require reports about the

[196] Google, Frontier Safety Framework: Version 3.0 6 (Sep. 22, 2025), https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/strengthening-our-frontier-safety-framework/frontier-safety-framework_3.pdf.
[197] Google, Frontier Safety Framework: Version 2.0, *supra* note 185, at 7.
[198] *Id.* at 3.
[199] *See* note 33-34 and accompanying text. *See also Responsibility & Safety*, Google DeepMind (July 21, 2025), https://deepmind.google/about/responsibility-safety/ (stating that the AGI Safety Council is led by Co-Founder and Chief AGI Scientist Shane Legg, and the Internal Responsibility and Safety Council is described as an "internal review group" co-chaired by the COO and VP of responsibility); *Alphabet Notice of 2025 Annual Meeting of Stockholders and Proxy Statement*, 23 (disclosing that the President of Global Affairs and Chief Legal Officer and Secretary, Alphabet and Google presides the Trust & Compliance Council).
[200] OpenAI, Preparedness Framework: Version 2 1 (Apr. 15, 2025), https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbddebcd/preparedness-framework-v2.pdf (last visited Aug. 14, 2025).
[201] *Id.* at 8.
[202] Sam Coggins et al., *The 2025 OpenAI Preparedness Framework Does Not Guarantee Any AI Risk Mitigation Practices: A Proof-of-Concept for Affordance Analyses of AI Safety Policies* (Oct. 13, 2025) (unpublished manuscript), http://arxiv.org/abs/2509.24394.
[203] OpenAI, Preparedness Framework: Version 2, *supra* note 200, at 15.

framework's application. Employees can also anonymously flag compliance concerns,[204] but ultimately, only the Board can reverse or revise course post facto.[205] The resulting governance structure is thus more advisory than binding: the Board and the CEO are in charge. As such, the framework remains subordinated to other considerations, namely competitive pressures. Indeed, the blog post announcing the updated policy flags that OpenAI will "adjust" (read: downgrade) safety requirements if a competitor releases a product without similar safeguards.[206]

3. Anthropic: Dialed Back Precautionary Approach

A leader in AI safety, Anthropic walked back self-governance commitments in February 2026. Version 2.2 of its Responsible Scaling Policy represents the apex of voluntary safety policies. Citing competitive pressures, the company relaxed its approach in the subsequent 3.0 version.

RSP v. 2.2: a Precautionary Approach

The May 2025 Responsible Scaling Policy represents the high watermark of voluntary safety measures. Anticipating future capabilities, it established safety levels ahead of time, with obligations corresponding to the increased risk level. The policy bound the company to safety measures ahead of time, committing the company to pause development until it could prove safety. The policy was precautionary in two important ways. In doubt about the risk level, it applied the guardrails of the higher-risk level.[207] Second, it committed to restricting development and deployment unless it could affirmatively prove safety.[208] The company thus assumed the burden and cost of proving safety, reminiscent of the "better safe than sorry" precautionary approach requiring the proponent of an uncertain technology with potentially serious or irreversible consequences to prove its safety prior to its roll-out.[209] As a reminder, the policy covers a narrow set of mostly hypothetical risks, such that its impact should not be overstated.[210] But within the narrow scope of catastrophic risks, the policy nevertheless bears the hallmarks of a precautionary approach.

In terms of governance, the policy actively involved the Board of Directors and Long-Term Benefit Trust in decisions to move forward. These entities were kept apprised of internal decisions to deploy and

[204] *Id.* at 12; *OpenAI's Raising Concerns Policy*, OpenAI, Jan. 12, 2026, https://openai.com/index/openai-raising-concerns-policy/ (last visited Aug. 26, 2025).
[205] OpenAI, Preparedness Framework: Version 2, *supra* note 200, at 3, 15.
[206] *Our updated Preparedness Framework*, OpenAI (Apr. 15, 2015), https://openai.com/index/updating-our-preparedness-framework/; Effie Webb, *OpenAI just gave itself wiggle room on safety if rivals release "high-risk" models*, Bus. Insider (Apr. 16, 2025), https://www.businessinsider.com/openai-safety-policy-gpt4-1-employee-criticism-musk-lawsuit-2025-4 (last visited Aug. 26, 2025).
[207] Responsible Scaling Policy - v 2.2, *supra* note 184, at § 3.3 (if the company cannot prove the system is below a risk threshold, it considers it above it).
[208] Responsible Scaling Policy - v 2.2, *supra* note 184, at § 4.3 (no further model development and deployment unless an affirmative safety case is established).
[209] For an example of this approach in the environmental context, see *Wingspread Statement on the Precautionary Principle*, https://www.gdrc.org/u-gov/precaution-3.html (last visited Feb. 15, 2024); Rio Declaration Principle 15; for a fuller discussion of precaution in the AI context, *see* Lévesque, *supra* note 120, at 110–111.
[210] *See supra* notes 188–90 and accompanying text.

further train models prior to their implementation.[211] This timing was crucial since reversing course is costlier and sometimes unfeasible. For example, if malicious actors breach security measures to exfiltrate model weights, attempts to de-deploy the model once it's widely available would be futile. Though the policy stopped short of giving the Board and the Trust a formal veto over decisions to push the frontier, its information-forcing mechanism fostered accountability towards those tasked with strategy and the long-term view within the company. The company further committed to a third-party review to assess compliance with procedural commitments in the scaling policy, but public-facing details about this process remain scant.

RSP v. 3.0: Softening Commitments and Governance

In February 2026, the company walked back the most stringent components of its scaling policy. It assessed that it could no longer afford self-imposed constraints in light of competitive pressure, the "anti-regulatory climate" and the federal government's pivot to economic growth.[212] With safety and regulation taking a backseat, the new policy focused on anchoring Anthropic's practices to industry trends. The new policy recommends that all frontier labs make a "strong argument" about safety.[213] Even on this lower standard, Anthropic warns that the company makes no guarantees regarding risk pertaining to non-novel chemical and biological weapons production and automated R&D. This contrasts with the previous 2.0 policy operating under the assumption of a higher risk unless proven otherwise and requiring an affirmative safety case before moving forward. The new policy replaces that precautionary approach with safety roadmaps, hoping that sharing them publicly will have a "forcing function" to prioritize safety among the company's multiple objectives such as immediate competition and commercial objectives.[214] In other words, the company is relying on internal name and shame to have the safety mission prevail over financial considerations.

Governance-wise, the new policy dulls oversight. The decision to move forward with new development and deployment now lies solely with the Responsible Scaling Officer, whereas it was previously shared with the CEO.[215] Moreover, the Trust and the Board are no longer privy to these decisions ahead of time. Though the revised policy retains safeguards like third-party audits to ensure that the company abides by the procedure it set out and protections for employees denouncing non-compliance,[216] overall it substantially rolls back previous commitments.

Having pioneered safety frameworks, Anthropic set the tone for frontier labs. In light of competitive pressures which show no sign of abating, especially as the prospect of federal regulation leveling the playing field dwindles, Anthropic downgraded to commitments it could realistically uphold. This

[211] Responsible Scaling Policy - v 2.2, *supra* note 184, at 7, § 3.3(4), 11, § 4.3(4) (the decision to proceed with deployment is shared with the Board and the Trust).
[212] *Responsible Scaling Policy Version 3.0*, (Feb. 24, 2026), https://www.anthropic.com/news/responsible-scaling-policy-v3.
[213] Anthropic's Responsible Scaling Policy (Version 3.0) § 1 (pp. 6-9) (2026), https://anthropic.com/responsible-scaling-policy/rsp-v3-0 [https://perma.cc/8RT2-F2WN].
[214] *Id.* § 2.
[215] *Id.* § 4(1).
[216] *Id.* §§ 4(3)-(4) (whistleblower protection and restrictions on non-disparagement clauses in employment contracts).

pragmatic decision is arguably preferable to letting the distance grow between commitments and actions that are no longer tenable, but nevertheless represents a significant degradation of scaling policies.

4. Business Structures Do Not Correlate With Differences Among Scaling Policies

Corporate structures do not meaningfully influence scaling policies. Although Anthropic's previous precautionary approach and proactive oversight stood out, the revised policy substantially diluted these commitments so as to erase any meaningful differences among leading labs. This suggests that when faced with competitive pressures, a public benefit corporation with a trust is no more conducive to carrying out its public interest safety commitments as compared to the traditional corporate business structure of Google or the OpenAI hybrid (a capped profit joint venture at the time of its latest framework). Ultimately, the scaling policies of all the companies under review fall short of meaningful guarantees, leaving significant gaps between these voluntary constraints and stated public interest missions.

Future research should consider more aspects of these scaling policies to surface other signals pertaining to the ways companies translate their mission into actionable scaling policies. Building on existing comparisons among scaling policies and deeper, company-specific dives,[217] future work should evaluate other factors like the scope of covered risks, the specificity of implementation requirements and the involvement of third parties at various stages of the policy application, as well as policies detailing compliance with specific laws.[218] Just to give a few examples of how other factors could modulate the present assessment, OpenAI and Google consider a broader array of capabilities than Anthropic, faring better in terms of the scope of covered dangerous capabilities.[219] Google also stands out for integrating outside expertise in evaluating capability thresholds and safeguards.[220] Similarly, its

[217] For examples of more comprehensive reviews of scaling policies, *see 2025 AI Safety Index*, Future of Life Institute, https://futureoflife.org/ai-safety-index-summer-2025/ (last visited Aug. 14, 2025) (grading OpenAI C, Anthropic C+ and GoogleDeepMind C- overall); *Comparison*, Risk Management Ratings, https://tracker.safer-ai.org/comparison/ (last visited Aug. 14, 2025); *Methodology*, Risk Management Ratings, https://tracker.safer-ai.org/methodology/ (last visited Aug. 14, 2025); *[Public] Do Companies' "AI Safety Policies" Meet Government Best Practice - Spreadsheet - Google Sheets*, https://web.archive.org/web/20240307115022/https://docs.google.com/spreadsheets/d/1G7Zuyfz40xiIz_A8HyORb7r-3SuV9bFEoPAeFBCMBB8/edit#gid=0 (last visited Aug. 14, 2025); *see also* Jide Alaga, Jonas Schuett & Markus Anderljung, *A Grading Rubric for AI Safety Frameworks* (Sep. 13, 2024) (unpublished manuscript), http://arxiv.org/abs/2409.08751; *Responsible Scaling: Comparing Government Guidance and Company Policy*, Bill Anderson-Samways, Responsible Scaling: Comparing Government Guidance and Company Policy (Mar. 11, 2024), https://www.iaps.ai/research/responsible-scaling (last visited Aug. 14, 2025) (evaluating Anthropic's Responsible Scaling Policy against best practices for responsible capability scaling issued by the UK Department for Science, Innovation and Technology in October 2023).

[218] OpenAI, Frontier Governance Framework (2026), https://cdn.openai.com/pdf/e37d949b-8c9f-4d76-b99e-4272f4631a7e/openai-frontier-governance-framework.pdf [https://perma.cc/7kb2-5vpc] (operationalizing safety for the EU AIA and the California Transparency in Frontier AI Act).

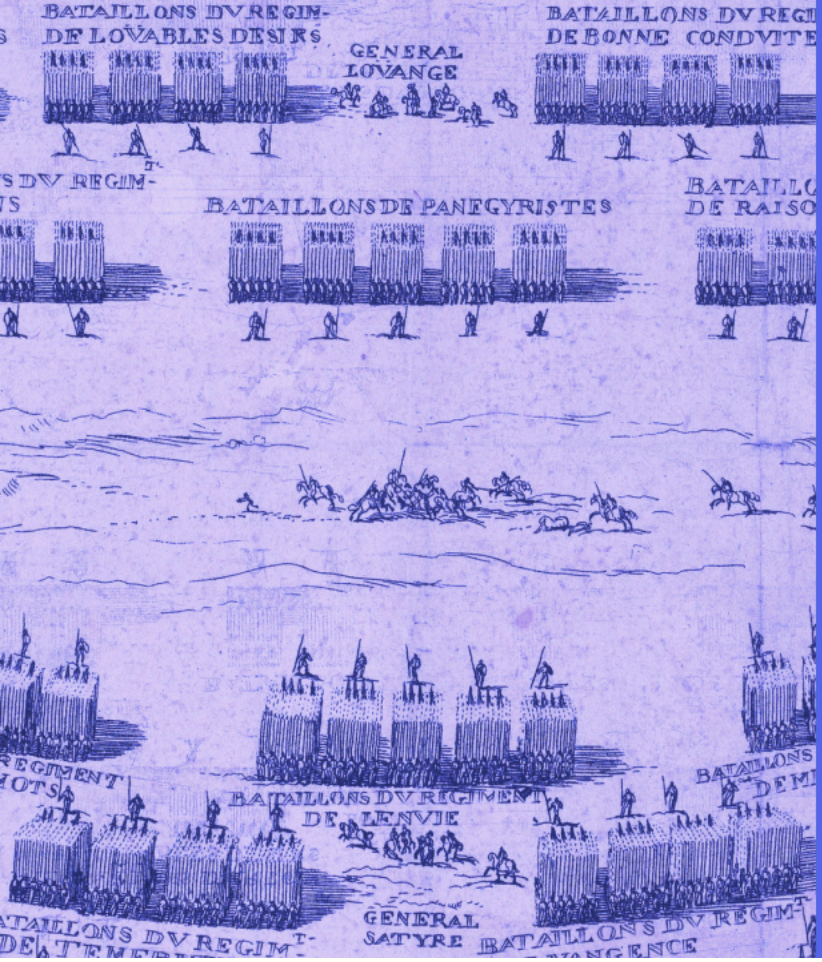
[219] OpenAI, Preparedness Framework: Version 2, *supra* note 200, at 4, 7 (in addition to tracking biochemical weapons, cyber capabilities and self-improvement, it monitors long range autonomy, "sandbagging" or the ability to thwart evaluations, replication and adaptation and undermining safeguards); Google, Frontier Safety Framework: Version 2.0, *supra* note 197 (tracking deception).

[220] Shevlane et al., *supra* note 195 (evaluation method is a collaboration among scientists of various research and academic institutions) (cited in Shevlane, *supra* note 194 and in Google, Frontier Safety Framework: Version 2.0, *supra* note 185, at 3).

safeguards incorporate externally designed security controls.[221] Future research should also consider information-sharing through the Frontier Model Forum, and how high-level commitments to "pace the frontier" stack against actual model release.[222]

[221] Google, Frontier Safety Framework: Version 3.0, *supra* note 196, at 8 (incorporating RAND security protocols); *see also* Nevo et al., *supra* note 187.

[222] Frontier Model Forum, https://www.frontiermodelforum.org/ (last visited Oct. 9, 2026); Dario Amodei, *We Must Pace the Frontier*, Dario Amodei (2026), https://darioamodei.com/post/we-must-pace-the-frontier; Dario Amodei (@DarioAmodei), X (Sep. 12, 2026, at 10:01 ET), https://x.com/DarioAmodei/status/2098773920774074715 (announcing the Pacing the Frontier essay on Sep. 12, suggesting it urged pacing 10 days before the release of a new model); *Introducing Claude Opus 5.5*, Anthropic (Sep. 22, 2026), https://www.anthropic.com/claude-opus-5-5 (reporting that the company's own benchmark measures a 10% lead on Fable 5.1 in agentic coding, a statistically significant advance even accounting for the standard error rate; and that the new model's improved performance also makes the frontier more accessible with broader diffusion and distillation possibilities); *see also* System Card: Claude Opus 5.5, Anthropic (Sep. 22, 2026), https://www-cdn.anthropic.com/fc1b44717c85dc068bc6ba5024219938094694bd/Claude%20Opus%205.5%20System%20Card.pdf (detailing Opus 5.5 capabilities and safeguards).



Military AI

Defense contracts are a good proxy for company self-interest. Not only are they lucrative—the Pentagon recently awarded contracts up to $200 million to each company under review—they also allow companies to stay in the good graces of regulators, presumably less inclined to impose constraints on strategic partners.[223] However, these contracts can collide with the mission of AI companies. Debates about the potential and benefits of military AI are divisive and context-sensitive, but at minimum raise important questions about whether such applications benefit humanity in line with the stated mission of AI companies.

Militaries have recently integrated AI into operations in two active theaters: Ukraine and the Middle East.[224] It is useful to disaggregate these deployments into lethal autonomous weapon systems (LAWS), AI-assisted weapons, and AI-based decision support systems (AI-DSS) involved in broader military

---

[223] Frank Konkel, *Pentagon Awards Multiple Companies $200M Contracts for AI Tools*, Nextgov (July 14, 2025), https://www.nextgov.com/acquisition/2025/07/pentagon-awards-multiple-companies-200m-contracts-ai-tools/406698/.

[224] Samuel Bendett, *Roles and Implications of AI in the Russian-Ukrainian Conflict*, CNAS (July 20, 2023), https://www.cnas.org/publications/commentary/roles-and-implications-of-ai-in-the-russian-ukrainian-conflict (last visited July 22, 2025); Emmanuel Grynszpan, *Russia Steps up Use of New Military Technologies against Ukraine, Including AI*, June 7, 2025, https://www.lemonde.fr/en/international/article/2025/06/07/russia-steps-up-use-of-new-military-technologies-against-ukraine-including-ai_6742108_4.html; Vitaliy Goncharuk, *Russia's War in Ukraine: Artificial Intelligence in Defence of Ukraine*, ICDS, Sep. 27, 2024, https://icds.ee/en/russias-war-in-ukraine-artificial-intelligence-in-defence-of-ukraine/; Letter Dated 8 March 2021 from the Panel of Experts on Libya Established Pursuant to Resolution 1973 (2011) Addressed to the President of the Security Council (2021) ¶ 63; p. 148 (Reporting a Turkish 'fire, forget and find' autonomous drone deployed in Libya); Jon Harper, *US Military Deploys New JADC2 Capability to Middle East*, DefenseScoop (Apr. 3, 2024), https://defensescoop.com/2024/04/03/centcom-jadc2-deploy-minimum-viable-capability/; Dexter Filkins, *Is the U.S. Ready for the Next War?*, New Yorker (July 14, 2025), https://www.newyorker.com/magazine/2025/07/21/is-the-us-ready-for-the-next-war; Yuval Abraham, *'A Mass Assassination Factory': Inside Israel's Calculated Bombing of Gaza*, +972 Magazine (Nov. 30, 2023), https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/; Heidy Khlaaf, *Safety and War: Safety and Security Assurance of Military AI Systems*, AI Now Institute 2 (June 25, 2024), https://ainowinstitute.org/publications/safety-and-war-safety-and-security-assurance-of-military-ai-systems.

applications. LAWS essentially operate without a human in the loop. The UN Secretary General maintains that such systems are "politically unacceptable and morally repugnant."[225] The current policy of the Pentagon subjects LAWS to pre-deployment approvals by senior military leaders, who have to verify "appropriate levels of human judgment over the use of force," a broad and flexible standard that doesn't require manual human control.[226] The Secretary of Defense can change the current directive unilaterally, and has to notify Congress after the fact.[227] Beyond LAWS, weapons with a human in the loop and AI-based decision support systems (AI-DSS) play a broader role in military applications. AI-DSS enhance intelligence analysis, surveillance and targeting—for example, analyzing combat terrain—generating mission plans, adjusting the trajectory of weapons to real-time environmental factors and even suggesting strike targets.[228] The distinction between LAWS, other weapons and AI-DSS is not always clear-cut, especially because automation bias may reduce human involvement to rubber-stamping AI suggestions.[229]

To grossly oversimplify debates pertaining to military AI, some raise concerns about the error rates of these systems while others expect AI to reduce collateral damage thanks to more accurate situational awareness.[230] Yet others argue that in practice, AI's potential for technical precision gives cover to indiscriminate attacks.[231] What remains clear is that military applications have very controversial implications in terms of these companies' stated mission to benefit humanity. Consequently, the

[225] *Lethal Autonomous Weapon Systems*, https://disarmament.unoda.org/en/our-work/emerging-challenges/lethal-autonomous-weapon-systems (last visited June 2, 2026). For a survey of LAWS, see Taylor Jones, *Real-Life Technologies That Prove Autonomous Weapons Are Already Here*, Future of Life Inst. (Nov. 22, 2021), https://futureoflife.org/aws/real-life-technologies-that-prove-autonomous-weapons-are-already-here/.

[226] U.S. Dep't of Def., Directive 3000.09, Autonomy in Weapon Systems (2023), https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf; Kelley M. Sayler, Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems (2026), https://www.congress.gov/crs-product/IF11150.

[227] Kelley M. Sayler, *supra* note 226.

[228] For a fuller mapping of AI military capabilities, see Sarah Grand-Clément, *Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain*, 12–13 (2023), https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/; *see also* Chris Panella, *This Ukrainian Tech Company Is Working to Beat Russia's Electronic Warfare without Hard-Wiring Drones to an Operator*, Bus. Insider (Feb. 20, 2025), https://www.businessinsider.com/ukrainian-company-could-have-new-answer-to-russias-drone-jamming-2025-2 (last visited July 22, 2025); Gulsanna Mamediieva, *Military AI: Lessons from Ukraine*, Tech Policy Press (Mar. 20, 2025), https://techpolicy.press/military-ai-lessons-from-ukraine; Patrick Tucker, *How AI Could Predict The Damage to Ukraine from Russian Missiles*, Defense One (Jan. 10, 2023), https://www.defenseone.com/technology/2023/01/how-ai-could-predict-damage-ukraine-russian-missiles/381633/; Charukeshi Bhatt & Tejas Bharadwaj, *Understanding the Global Debate on Lethal Autonomous Weapons Systems: An Indian Perspective*, Carnegie Endowment for International Peace, https://carnegieendowment.org/research/2024/08/understanding-the-global-debate-on-lethal-autonomous-weapons-systems-an-indian-perspective?lang=en (last visited July 22, 2025).

[229] Jessica Dorsey, *The erosion of human(e) judgement in targeting? Quantification logics, AI-enabled decision support systems and proportionality assessments in IHL*, Int'l Rev. Red Cross (Dec. 1, 2025), http://international-review.icrc.org/articles/the-erosion-of-humane-judgement-in-targeting-quantification-logics-ai-enabled-decision-support-systems-and-proportionality-assessments-in-ihl-930 (noting the risk of rubber-stamping AI-DSS suggestions).

[230] Yuval Abraham, *'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza*, +972 Magazine (Apr. 3, 2024), https://www.972mag.com/lavender-ai-israeli-army-gaza/ [https://perma.cc/5APC-E3L3] (reporting that human personnel spends 20 seconds reviewing an AI-generated target despite a 10% error rate); United States of America, *Humanitarian benefits of emerging technologies in the area of lethal autonomous weapon systems*, U.N. Doc. CCW/GGE.1/2018/WP.4 (Apr. 3, 2018) (arguing that AI can lift the fog of war such that it can identify military targets with more precision).

[231] Abraham, *Lavender, supra* note 230 (reporting on AI generating 'power targets' at high speeds, including private residences, public buildings and high-rise blocks intended to shock the civilian population into pressuring terrorists); Filkins, *supra* note 224 (observer noting that A.I. lends a veneer of legitimacy to a policy of complete destruction); Harry Davies, Bethan McKernan & Dan Sabbagh, '*The Gospel': How Israel Uses AI to Select Bombing Targets in Gaza,* Guardian (Dec. 1, 2023), https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets.

decision to engage in defense contracts provides a useful lens to analyze how companies navigate potential tensions between their mission and revenue-generating activities. The case studies below show that irrespective of its business structure, every company under review has defense contracts.

In July 2025, the Pentagon enlisted all three companies to "develop prototype frontier AI capabilities to address critical national security challenges in both war fighting and enterprise domains," each awarded up to $200 million to deliver such services.[232] The Pentagon's press release described how it intends to use frontier AI to develop agentic and generative AI capabilities, with the Chief Digital Officer stating that "[l]everaging commercially available solutions into an integrated capabilities approach will accelerate the use of advanced AI as part of our Joint mission essential tasks in our warfighting domain as well as intelligence, business, and enterprise information systems."[233] This portends a more prominent role for AI in the battlefield as the U.S. is moving towards autonomous weapons. For example, in August 2023 the Deputy Secretary of Defense described the Replicator Initiative to build thousands of autonomous ships and aircraft "of various sizes and payloads" in the next two years.[234] But even before these contracts, the companies under review were involved in military activities through partnerships with defense contractors or through cloud services to government clients.

1. Google: Reversing Course on Weapons and Surveillance Applications

DeepMind and Google substantially watered down their policy over time, and did not always act consistently with said policies. For DeepMind, a ban on all military applications was a precondition to the Google acquisition. Its founder stated that "[o]ne constraint we do have— […] part of the acquisition terms—is that no technology coming out of DeepMind will be used for military or intelligence purposes."[235] Although the research lab was loud and clear about such restrictions being a

[232] *CDAO Announces Partnerships with Frontier AI Companies to Address National Security Mission*, Chief Digital and Artificial Intelligence Office, https://www.ai.mil/Latest/News-Press/PR-View/Article/4242822/cdao-announces-partnerships-with-frontier-ai-companies-to-address-national-secu/ [https://web.archive.org/web/20260712165210/https://www.ai.mil/Latest/News-Press/PR-View/Article/4242822/cdao-announces-partnerships-with-frontier-ai-companies-to-address-national-secu/] (last visited July 22, 2025).

[233] *CDAO Announces Partnerships with Frontier AI Companies to Address National Security Mission*, *supra* note 232.

[234] *Deputy Secretary of Defense Hicks Announces First Tranche of Replicator Capabilities Focus*, U.S. Dep't of Def., https://www.defense.gov/News/Releases/Release/Article/3765644/deputy-secretary-of-defense-hicks-announces-first-tranche-of-replicator-capabil/ (last visited Aug. 29, 2025); *The Replicator Initiative*, https://www.diu.mil/replicator [https://web.archive.org/web/20250729192835/https://www.diu.mil/replicator] (last visited July 24, 2025); *Deputy Secretary of Defense Kathleen Hicks' Remarks: "Unpacking the Replicator Initiative,"* U.S. Dep't of Def. (Sep. 6, 2023), https://www.defense.gov/News/Speeches/Speech/Article/3517213/deputy-secretary-of-defense-kathleen-hicks-remarks-unpacking-the-replicator-ini/; Kathleen Hicks, Deputy Secretary, Dep't of Def., Keynote Address at the NDIA Emerging Technologies for Defense Conference: The Urgency to Innovate, U.S. Dep't of Def. (Aug. 28, 2023), https://www.defense.gov/News/Speeches/Speech/Article/3507156/deputy-secretary-of-defense-kathleen-hicks-keynote-address-the-urgency-to-innov/; Kyle Hiebert, *The United States Quietly Kick-Starts the Autonomous Weapons Era*, Ctr. for Int'l Governance Innovation (Jan. 15, 2024), https://www.cigionline.org/articles/the-united-states-quietly-kick-starts-the-autonomous-weapons-era/ (last visited July 22, 2025); *see also U.S. 5th Fleet Conducts Unmanned Exercise in the Arabian Gulf*, DVIDS (2023), https://www.dvidshub.net/video/902487/us-5th-fleet-conducts-unmanned-exercise-arabian-gulf (military exercise launching attack from unmanned ship).

[235] Levy, *supra* note 15.

key condition of its acquisition in 2015,[236] they got lost in the merger with the revenue-focused Google division of Alphabet.[237]

The company faced significant backlash over Project Maven, an AI-DSS that could accelerate the "kill chain" by suggesting strike targets.[238] The U.S. initiative sought to analyze massive amounts of data, including drone footage, to inform target acquisition. Google did not renew its contract with the Pentagon and committed to principles foregoing applications likely to cause overall harm and weapons, but continued working with the military on other areas.[239] Reporting indeed suggests that the company offered AI-DSS services to other militaries via cloud computing services. Cloud video intelligence was available to a military subcontractor,[240] and statements by government officials indicate that Google Cloud supported combat operations,[241] raising concerns about complicity with human rights violations.[242] Google's cloud services provided a gateway to serve AI models to governments, including military actors.[243]

In February 2025, Google lifted the ban on weapons and surveillance applications and dropped the reference to harmful applications, reasoning that "companies, governments, and organizations sharing these values should work together to create AI that protects people, promotes global growth, and supports national security."[244] Shortly thereafter in July 2025, the Pentagon enlisted Google as a military contractor. Insofar as Google's corporate objective is to make money for its shareholders, pursuing military contracts is consistent with said objective. Whether such contracts comport with the

---

[236] *Id.*
[237] *See supra* Part I, "A Cog in the Alphabet Machine."
[238] PBS, *Amid Pressure from Employees, Google Drops Pentagon's Project Maven Account*, PBS NewsHour (June 3, 2018), https://www.pbs.org/newshour/show/amid-pressure-from-employees-google-drops-pentagons-project-maven-account; Kate Conger, *Google Employees Resign in Protest Against Pentagon Contract*, Gizmodo (May 14, 2018), https://gizmodo.com/google-employees-resign-in-protest-against-pentagon-con-1825729300 [https://web.archive.org/web/20250106220519/https://gizmodo.com/google-employees-resign-in-protest-against-pentagon-con-1825729300].
[239] Pichai, *AI at Google*, *supra* note 4.
[240] Gerrit De Vynck, *Google Helped Israeli Military Contractor with AI, Whistleblower Alleges*, Wash. Post (Feb. 1, 2026), https://www.washingtonpost.com/technology/2026/02/01/google-ai-israel-military/.
[241] Yossi Hatoni, *Ein Lanu Breira — Ha-Anan Hu Ha-Atid* [*We Have No Choice — The Cloud Is the Future*], Anashim u-Mahshevim [People & Computers] (Feb. 29, 2024), https://www.pc.co.il/featured/404442/ (quoting the head of Israel's National Cyber Directorate: "Due to the Nimbus public cloud, phenomenal things happen in combat, which constitute a significant part of victory — and I won't go into detail") (author's translation, AI-assisted); Gerrit De Vynck, *Google Rushed to Sell AI Tools to Israel's Military After Hamas Attack*, Wash. Post (Jan. 21, 2025), https://www.washingtonpost.com/technology/2025/01/21/google-ai-israel-war-hamas-attack-gaza/.
[242] *Biddle*, *supra* note 36.
[243] Sam Biddle, *Documents Reveal Advanced AI Tools Google Is Selling to Israel*, Intercept (July 24, 2022), https://theintercept.com/2022/07/24/google-israel-artificial-intelligence-project-nimbus/ (reporting that the training material cites Ministry of Defense as an example); *Google Cloud Core Infrastructure* 148 (2021), https://www.documentcloud.org/documents/22119705-core-*infra*-nimbus-webinar/ (showing facial recognition and sentiment analysis capabilities available to government client); Google Cloud, *Welcome to Cloud OnBoard: Big Data and Machine Learning* 106 (July 24, 2022), https://www.documentcloud.org/documents/22119704-webinar-big-data-ml/ (describing object tracking capabilities); *see also* De Vynck, *supra* note 241 (reporting that Google employees requested access to Gemini for military cloud clients to process audio and video); Caroline Haskins, *The Hidden Ties Between Google and Amazon's Project Nimbus and Israel's Military*, Wired (July 15, 2024), https://www.wired.com/story/amazon-google-project-nimbus-israel-idf/ (last visited July 24, 2025); Uri Berkovitz, *Proyekt ha-Anan ha-Anak shel ha-Memshala: Ha-Krav me-Ahorei ha-Kla'im al Nimbus* [*The Government's Giant Cloud Project: The Behind-the-Scenes Battle over Nimbus*], N12 (Jan. 24, 2022), https://www.mako.co.il/news-business/news/Article-46ecb25d3078e71027.htm (describing the IDF as the original client for Project Nimbus) (author's translation, AI-assisted).
[244] *AI Principles*, Google AI, *supra* note 33; *Responsible AI: Our 2024 Report and Ongoing Work*, Google (Feb. 4, 2025), https://blog.google/technology/ai/responsible-ai-2024-report-ongoing-work/; Dave, *supra* note 33; Lucy Hooker & Chris Wallace, *Google Ending AI Arms Ban Incredibly Concerning, Campaigners Say*, BBC (Feb. 5, 2025), https://www.bbc.com/news/articles/cy081ngx2zjo.

stated mission to "responsibly develop and maintain advanced AI for the long-term benefit of humanity" is less clear.

2. OpenAI: From Banning Military Applications to Defense Contractor

OpenAI initially opposed all military deployments, but gradually relaxed its position. Up until January 2024, it specifically prohibited military and warfare applications in addition to weapons development.[245] It then removed prohibitions on military and warfare applications, leaving in place prohibitions against weapon use and development.[246] But even before the official policy reversal, it began offering AI capabilities to the U.S. military via its partnership with Microsoft, which served OpenAI models to its Azure cloud clients. At first, the cloud platform only offered OpenAI's models to commercial clients. In June 2023, Azure extended access to OpenAI models for government clients.[247] A Microsoft presentation suggested the Pentagon use OpenAI's Dall-E model to generate images for battle management training.[248] This calls into question public-facing policies prohibiting military applications prior to the policy change in January 2024.

After its policy change allowing military applications, OpenAI increased touch points between its technology and the battlefield. It partnered with U.S. defense contractor Anduril to assess drone threats in December 2024,[249] then secured its first direct defense contract in July 2025.[250] OpenAI initially framed this contract as bringing AI to the government in service of the public good.[251] It further claimed that the contract leveraged frontier AI for "administrative operations," noting that usage policies and guidelines against weapons development still applied.[252] Reiterating the company's policy against using its products for the development and use of weapons, an OpenAI spokesperson framed

---

[245] Hayden Field, *OpenAI Quietly Removes Ban on Military Use of Its AI Tools*, CNBC (Jan. 16, 2024), https://www.cnbc.com/2024/01/16/openai-quietly-removes-ban-on-military-use-of-its-ai-tools.html; *Usage Policies*, (Mar. 23, 2023), https://web.archive.org/web/20240109122522/https:/openai.com/policies/usage-policies ("We don't allow the use of our models for […] [a]ctivity that has high risk of physical harm, including […] [w]eapons development [and] [m]ilitary and warfare").
[246] *Usage Policies*, https://openai.com/policies/usage-policies/ (last visited July 22, 2025).
[247] Bill Chappell, *Azure OpenAI Service: Transforming Workloads for Azure Government*, Microsoft Azure Blog (June 7, 2023), https://azure.microsoft.com/en-us/blog/unlock-new-insights-with-azure-openai-service-for-government/; Rachel Metz, *Microsoft Is Bringing OpenAI's GPT-4 AI Model to US Government Agencies*, Bloomberg (June 7, 2023), https://www.bloomberg.com/news/articles/2023-06-07/microsoft-offers-powerful-openai-technology-to-us-government-cloud-customers (confirming that the DoD will be experimenting with OpenAI's models through Azure).
[248] Nehemiah Kuhns, *Generative AI with DoD Data (2023)*, https://www.documentcloud.org/documents/24538175-generative-ai-with-dod-data_microsoft/ [https://perma.cc/63SL-CQY4]; the presentation date could not be independently verified, but is reported to be October 17, 2023 in Jack Poulson, *'Your AI is Your Rifle': Leak sheds light on ecosystem behind Pentagon's AI adoption*, All-Source Intel. (Apr. 10, 2024), https://jackpoulson.substack.com/p/your-ai-is-your-rifle; Sam Biddle, *Microsoft Pitched OpenAI's DALL-E as Battlefield Tool for U.S. Military*, Intercept (Apr. 10, 2024), https://theintercept.com/2024/04/10/microsoft-openai-dalle-ai-military-use/.
[249] *Anduril Partners with OpenAI to Advance U.S. Artificial Intelligence Leadership and Protect U.S. and Allied Forces*, https://www.anduril.com/anduril-partners-with-openai-to-advance-u-s-artificial-intelligence-leadership-and-protect-u-s/ (last visited July 23, 2025); Will Knight, *OpenAI Is Working With Anduril to Supply the US Military With AI*, Wired (Dec. 4, 2024), https://www.wired.com/story/openai-anduril-defense/ (last visited July 23, 2025) [https://www.anduril.com/news/anduril-partners-with-openai-to-advance-u-s-artificial-intelligence-leadership-and-protect-u-s].
[250] Usage policies, *supra* note 246; Sam Biddle, *OpenAI Quietly Deletes Ban on Using ChatGPT for "Military and Warfare,"* Intercept (Jan. 12, 2024), https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/; Agence France-Presse, *OpenAI Wins $200m Contract with US Military for 'Warfighting,'* Guardian (June 17, 2025), https://www.theguardian.com/technology/2025/jun/17/openai-military-contract-warfighting.
[251] *Introducing OpenAI for Government*, https://openai.com/global-affairs/introducing-openai-for-government/ (last visited July 22, 2025).
[252] *Id.*

the contract as exploring ways "OpenAI's tools can improve military operations and cybersecurity, save time for staff by making tedious work more efficient, and help the Pentagon better support service members."[253]

Yet in February 2026, the company seemed to relax its standards once again. Only hours after Anthropic refused to supply models to the Pentagon over concerns about unreliable autonomous weapons and invasive mass surveillance, the company stepped in. Facing a barrage of criticism for the move which, by the CEO's own admission, looked "opportunistic and sloppy,"[254] it added contractual and technical guardrails. Whether these guardrails effectively allow the company to control military uses and ultimately promote its mission remains uncertain at best. The contract enables the Pentagon to use OpenAI's products for all lawful purposes consistent with applicable laws, operational requirements and safety and oversight protocols and prohibits using those products to independently direct autonomous weapons where laws, regulations and policies require human control. However, the Pentagon can unilaterally amend these policies,[255] giving OpenAI little actual control over deployments. A leaked internal memo from Anthropic's CEO captures these concerns:

> "On autonomous weapons, the Pentagon claims that 'human in the loop is the law', but they are incorrect. It is currently Pentagon policy (set during the Biden admin) that a human has to be in the loop of firing a weapon. But that policy can be changed unilaterally by Pete Hegseth, which is exactly what we are worried about. So it is not, for all intents and purposes, a real constraint."[256]

OpenAI asserts that subsequent policy changes would not impact the obligations in the contract,[257] as it intended "applicable law" to mean "the law applicable at the time the contract is signed".[258] At minimum, the ambiguity between the plain reading of the contract meaning current law subject to change, and OpenAI's subsequent clarification that it intended to hardcode the obligations in place at the time of the agreement, suggests significant uncertainty.

On the technical side, the company cannot effectively verify that a human pre-approved a kill strike decision.[259] Furthermore, the company implies that preventing on-device computing (known as edge

---

[253] Jon Harper, *Pentagon Tapping OpenAI, Other Vendors for 'Frontier AI' Projects*, DefenseScoop (June 17, 2025), https://defensescoop.com/2025/06/17/pentagon-openai-frontier-ai-projects-cdao/.

[254] For claims that the deal has adequate guardrails, see Chris Vallance & Laura Cress, *OpenAI changes deal with US military after backlash*, BBC (Mar. 3, 2026), https://www.bbc.com/news/articles/c3rz1nd0egro; *Our Agreement with the Department of War*, OpenAI (May 18, 2026), https://openai.com/index/our-agreement-with-the-department-of-war/.

[255] Sigal Samuel, *The One Question Everyone Should Be Asking after OpenAI's Deal with the Pentagon*, Vox (Mar. 3, 2026), https://www.vox.com/future-perfect/481322/pentagon-anthropic-openai-surveillance-china; Alan Rozenshtein (@ARozenshtein), X (Feb. 28, 2026, at 17:04 ET), https://x.com/ARozenshtein/status/2027867469503951297.

[256] Stephanie Palazzolo, *Read Anthropic CEO's Memo Attacking OpenAI's 'Mendacious' Pentagon Announcement*, Information (Mar. 4, 2026), https://www.theinformation.com/articles/read-anthropic-ceos-memo-attacking-openais-mendacious-pentagon-announcement (reproducing Dario Amodei's internal memo responding to OpenAI's Pentagon announcement).

[257] *Our agreement with the Department of War*, *supra* note 254 ("Our contract explicitly references the surveillance and autonomous weapons laws and policies as they exist today, so that even if those laws or policies change in the future, use of our systems must still remain aligned with the current standards reflected in the agreement[.]") (emphasis omitted).

[258] NatSecKatrina (@natseckatrina), X (Feb. 28, 2026, at 19:48 ET), https://x.com/natseckatrina/status/2027908878952722693 [https://perma.cc/H73L-WUE8] ("[W]e intended it to mean 'the law applicable at the time the contract is signed.'").

[259] Hayden Field, *How OpenAI Caved to the Pentagon on AI Surveillance*, Verge (Mar. 2, 2026), https://www.theverge.com/ai-artificial-intelligence/887309/openai-anthropic-dod-military-pentagon-contract-sam-altman-hegseth.

computing) precludes autonomous weapons.[260] However, multiple pathways enable autonomous weapons to run on cloud architectures. Small servers deployed directly on the battlefield can already store and process data to support military operations.[261] As the technology progresses, these servers could communicate a strike decision to a remote weapon in milliseconds, effectively blurring the cloud/edge distinction. In any event, OpenAI's recent defense contracts indicate that it is willing to engage in controversial military activities, with tenuous guardrails far from assuaging concerns about autonomous weapons. Ultimately, OpenAI's military contracts call into question its ability to fulfill its stated fiduciary duty to humanity.

3. Anthropic: Military Contractor Upholding Usage Policies

Anthropic is deeply intertwined with the U.S. military, providing AI-DSS in multiple contexts to assist intelligence and war-fighting efforts. For example, it partnered with defense contractor Palantir in November 2024 to offer its flagship model Claude in classified intelligence and military environments. The nature of usage is steeped in abstract lingo like "enhancing analytical capabilities and operational efficiencies in vital government operations"[262] such that the exact applications of Anthropic models remain murky.

Unlike OpenAI, which dropped its explicit military and warfare ban in Jan 2024, Anthropic never had one to begin with. Its usage policies prohibit certain military applications, but enable the company to tailor contractual carve outs with government clients who show adequate safeguards,[263] effectively creating a discretionary permissive regime for military applications.

The Pentagon had awarded Anthropic the same $200 million defense contract as OpenAI, raising similar questions about The Pentagon complying with the company's usage policies. Tensions came to a head in February 2026. After the use of Anthropic's technology in the U.S. operation to capture Maduro in Venezuela (an operation carried on shaky legal grounds),[264] the company reportedly required assurances that its products would not be used to carry out mass surveillance on Americans or to develop fully autonomous weapons in violation of its usage policies.[265] The Pentagon opposed such restrictions, arguing that as long as the government complied with U.S. law it should have

---

[260] *Our Agreement with the Department of War*, OpenAI, *supra* note 254 ("We are not [. . .] deploying our models on edge devices (where there could be a possibility of usage for autonomous lethal weapons)[.]").
[261] *Microsoft Tactical Cloud Platform*, https://wwps.microsoft.com/wp-content/uploads/2023/12/Microsoft-Tactical-Cloud-Platform-Whitepaper.pdf (describing low-latency tactical edge compute).
[262] *Palantir IR*, (Nov. 7, 2024), https://investors.palantir.com/news-details/2024/Anthropic-and-Palantir-Partner-to-Bring-Claude-AI-Models-to-AWS-for-U.S.-Government-Intelligence-and-Defense-Operations/.
[263] *Usage Policy*, Anthropic (Sep. 15, 2025), https://www.anthropic.com/legal/aup; *Exceptions to Our Usage Policy*, Claude (Mar. 16, 2026), https://support.claude.com/en/articles/9528712-exceptions-to-our-usage-policy; *see also Usage Policy - Previous Version*, Anthropic (June 5, 2024), https://www.anthropic.com/legal/archive/3ce67c8d-e604-4e72-b2f2-6f060b08ddd6 (last visited July 23, 2025); *Acceptable Use Policy*, Anthropic (Sep. 15, 2023), https://www.anthropic.com/legal/archive/4903a61b-037c-4293-9996-88eb1908f0b2.
[264] *The US Capture of Venezuela's Maduro: An International Legal Analysis*, Univ. Cambridge (Jan. 8, 2026), https://www.cam.ac.uk/research/news/the-us-capture-of-venezuelas-maduro-an-international-legal-analysis.
[265] Justin Hendrix, *A Timeline of the Anthropic-Pentagon Dispute*, Tech Policy Press (Feb. 25, 2026), https://www.techpolicy.press/a-timeline-of-the-anthropic-pentagon-dispute/ (last visited June 3, 2026).

unconstrained access to commercial AI tools.[266] It dismissed Anthropic's concerns, and designated it a supply chain risk, a label normally reserved to blacklist foreign adversaries. Litigation is ongoing, with courts divided over the designation: a federal district court held it unlawful, while the D.C. Circuit upheld it.[267] Though the company has earned accolades for holding steady against its powerful government counterpart, it is worth bearing in mind that it opposed autonomous weapons not on principle, but rather because they were still unreliable and lacked proper oversight,[268] with the CEO stating that "[e]ven fully autonomous weapons […] may prove critical for our national defense."[269] What's more, the focus on fully autonomous weapons should not detract from the company's deep involvement with military AI-DSS offering threat identification and targeting suggestions, as well as partnerships with Palantir, problematic in their own right. Like all the other companies under review, Anthropic ultimately remains in the business of war.

4. Business Structures Have No Significant Effect on Defense Contracts

Military applications are a good litmus test for the commitment to prioritize humanity's benefit. In the abstract, providing AI technology might come across as patriotic and ethical, especially when it comes to a company's own domestic military. That said, the companies under review have often done so against their own internal policies. In that sense, they violate the mission as specified in internal policies. Moreover, Google and OpenAI have offered their products to U.S. allies whose military operations have raised significant concerns in terms of compliance with international law, raising the prospect of AI-enabled war crimes. Though controversies about AI's potential to improve warfighting are beyond the scope of this report, the practical deployments of AI thus far suggests that it enables rather than reduces indiscriminate attacks.

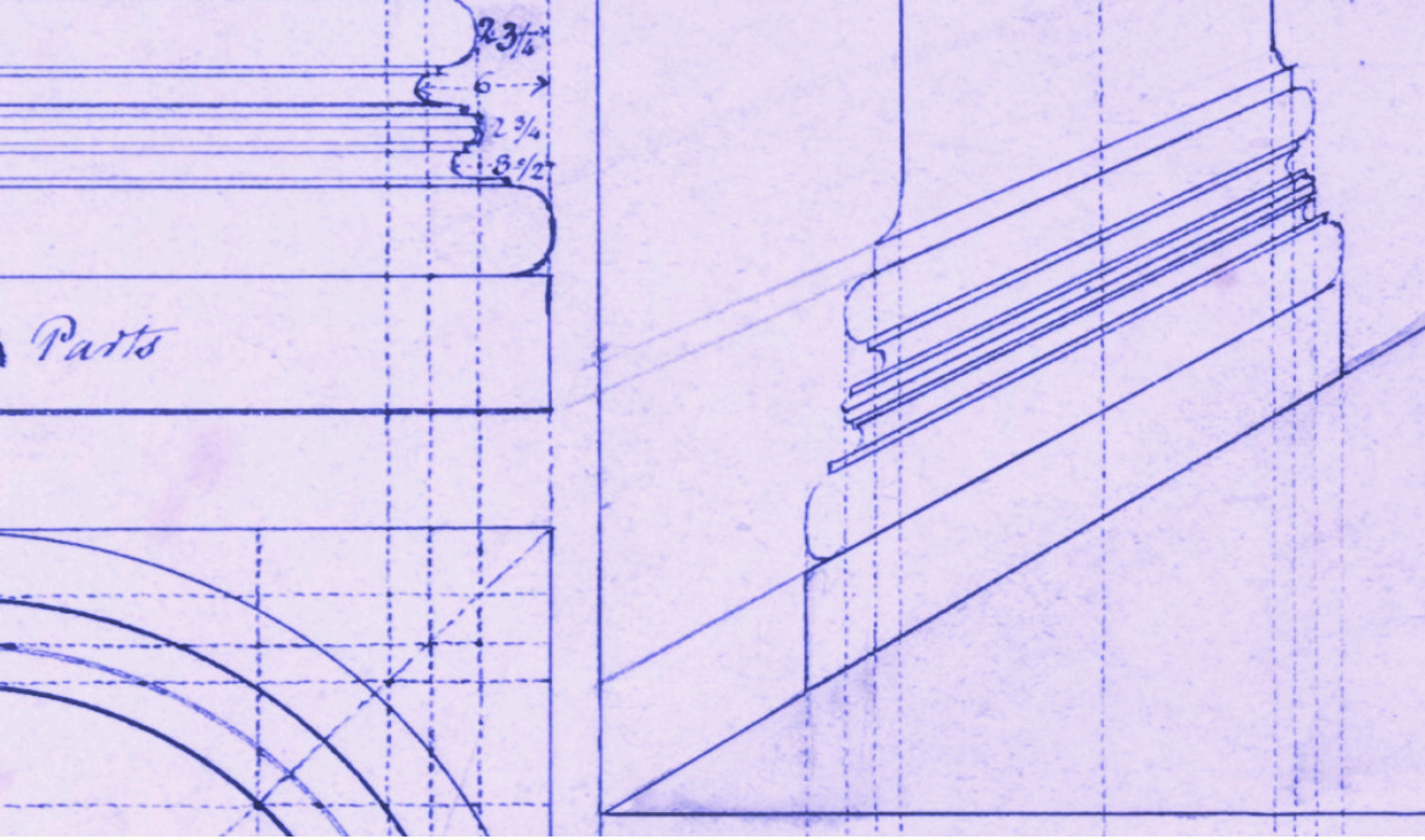
Every company under review serves their models to military actors, suggesting that none of the business structures under review can withstand the lucrative and strategic appeal of military contracts. Though usage policies tend to prohibit LAWS and other weapons, they still allow controversial AI-DSS applications. Whether such usage policies can effectively control military deployments remains an open question, as shown by the debate about OpenAI's contractual limitations of Pentagon uses. The evolution of military policy is also quite revealing, as it tracks changes in the structure of AI labs. For example, the erosion of DeepMind's broad commitment to forego military applications tracks its amalgamation deeper into the Google ecosystem (see Annex 1 for a timeline). Similarly, OpenAI dropped the ban of military applications during its transition from a non-profit to a capital-seeking "capped-profit" structure.

---

[266] Deepa Seetharaman et al., *Exclusive: Pentagon Clashes with Anthropic over Military AI Use, Sources Say*, Reuters (Jan. 29, 2026), https://www.reuters.com/business/pentagon-clashes-with-anthropic-over-military-ai-use-2026-01-29/.
[267] *Anthropic PBC v. U.S. Dep't of War*, No. 3:26-cv-01996 (N.D. Cal. Aug. 27, 2026) (holding the designation unlawful as First Amendment retaliation and beyond statutory authority); *Anthropic PBC v. U.S. Dep't of War*, No. 26-1049 (D.C. Cir. Sep. 25, 2026) (upholding the designation in a parallel challenge).
[268] *Statement from Dario Amodei on Our Discussions with the Department of War*, Anthropic (Feb. 26, 2026), https://www.anthropic.com/news/statement-department-of-war.
[269] *Id.*

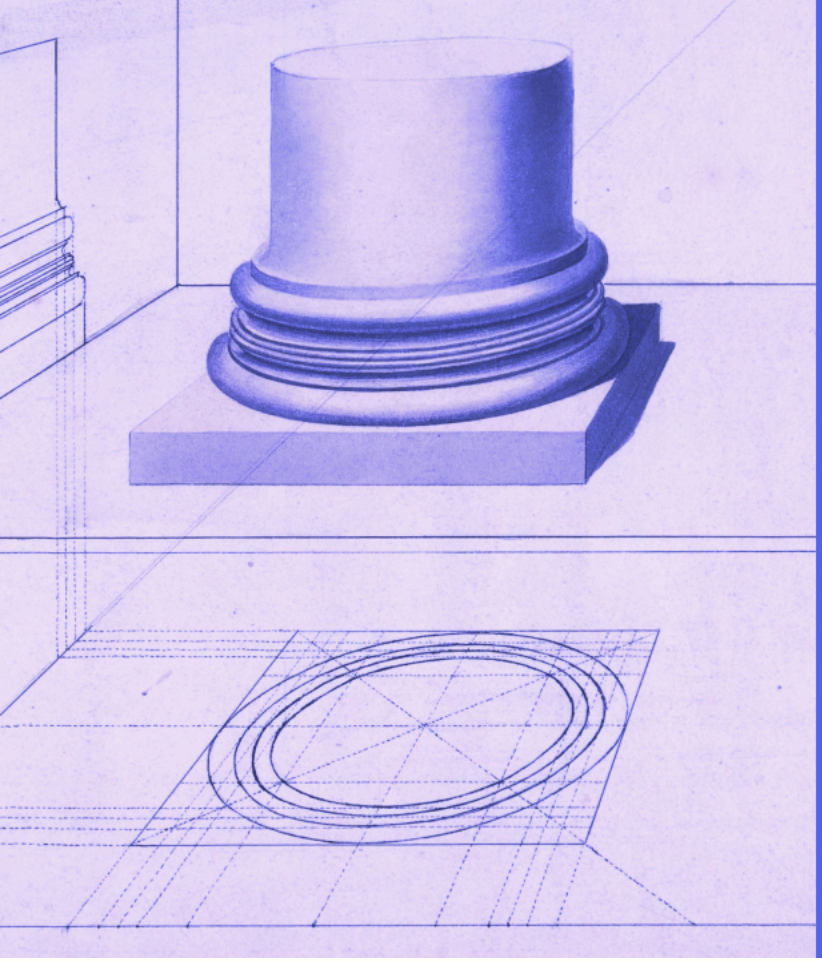
Overall, these case studies suggest that corporate structure matters to upholding the mission, but only to a certain extent. Even measured by imperfect yardsticks like regulatory, scaling and military policies, the most innovative company structures eventually buckle under financial pressure. Anthropic's combination of a public benefit corporation and a long-term benefit trust seems more responsive to the mission when it comes to scaling policies—until competitive pressures made unilateral commitments unsustainable. Remarkably, no company under review has maintained a principled stance against military contracts, implying that business structure alone cannot withstand the siren song of military contracts.

Put differently, creative corporate vehicles cannot protect the mission when that involves leaving money on the table. In fact, the point of these vehicles is arguably to allow, rather than require, companies to do so. No company can be expected to self-constrain when competitors don't. Indeed, both OpenAI and Anthropic have signaled that they will not let safety practices stand in the way of commercializing new products and fall behind while their competitors push the frontier. This collective action problem is an eloquent call for regulation if there ever was one. Similarly, defense contracts vividly illustrate that neither OpenAI's non-profit corporation nor Anthropic's public interest corporation, let alone Google's traditional for-profit corporation, have the right incentives to walk the talk on their own. That said, there are levers for corporate stakeholders and external actors to activate in order to hold AI companies to their word. The next section tools them to align corporate actions with their stated public interest mission



Part III. Levers to Enforce the Mission

Bespoke corporate structures have not always delivered in practice when it comes to prioritizing mission over profit. However, corporate governance offers levers to hold AI companies to their word. This section explores legal tools within the corporate and securities law toolkit, listing the main recourses for each type of corporate structure under review and assessing their potential and limit



General Shareholder Recourse and Mandatory Disclosures

Shareholders primarily want to make money from their stocks, so their interest in forcing the company to fulfill its mission at the expense of profits is limited. That said, some activist shareholders use their voice to influence a company's policy through a variety of strategies, some of which are examined below. Although shareholder recourses primarily concern Alphabet (Google's parent company), recourses extend to Anthropic PBC, and as of recently, OpenAI PBC. Moreover, the planned IPO would subject Anthropic to disclosure requirements as well.

1. Shareholder Recourses

Shareholders can contest management decisions inconsistent with the mission through formal resolutions. However, the founders of Google retain a controlling share of voting stock as outlined in Figure 5 above.[270] Therefore, shareholder resolutions have a limited impact on the company's activity. Still, shareholders have proposed resolutions related to AI governance in order to shape the company's activities. For example, a 2025 resolution sought greater due diligence regarding the use of Google products for surveillance, censorship, and military purposes.[271] Another pertained to the risk of discrimination of generative AI, and a third one raised concerns about the company's lobbying efforts

---

[270] Kamil Franek, *Who Really Owns Google (Alphabet) and Who Controls It,* Kamil Franek | Bus. Analytics (Mar. 1, 2023), https://www.kamilfranek.com/who-owns-google-alphabet/; *Google Rejects All 13 Shareholder Proposals at Its Annual Meeting, despite Protesting Workers*, Packt (June 20, 2019), https://www.packtpub.com/en-us/learning/how-to-tutorials/google-rejects-all-13-shareholder-proposals-at-its-annual-meeting-despite-protesting-workers (last visited Feb. 13, 2025); Alphabet, *supra* note 29, art. IV 2(a) (voting rights); Alphabet Bylaws, *supra* note 29 § 2. 9 (simple majority voting).

[271] *Alphabet Investor Relations - Investors - 2025 Annual Meeting of Stockholders*, https://abc.xyz/investor/annual-meeting/ (last visited Nov. 21, 2025).

to oppose child safety regulation.[272] Though none passed, they sent a signal that these issues are on the minds of some shareholders—besides controlling founders.

In theory, shareholders can also sue directors on behalf of the company for failure to uphold their fiduciary duties. Called a derivative action, this is the standard avenue to hold the board of directors accountable. But in practice, a derivative action against a board straying from the company's mission has next to no prospect of success. For a traditional profit-maximizing corporation like Alphabet, a derivative action is procedurally easy but substantively empty: any shareholder can sue on the corporation's behalf, but the mission is not a legally protected object of fiduciary duty. Consequently, straying from principles in pursuit of profit is precisely the conduct the business judgment rule shields.

In public benefit corporations like OpenAI and Anthropic, the reverse is true: substantively sustainable, but procedurally challenging. Section 365 of the Delaware code makes the public benefit a mandatory consideration for directors, giving mission-related claims a credible liability hook. However, § 367 reserves derivative enforcement to holders of 2% of shares (or $2 million in a publicly listed company). However, shareholders could band together to reach these thresholds, making coalition-building among like-minded shareholders an important strategic avenue for the future. In terms of liability, both companies exculpate directors from balancing failures as much as the law permits.[273] No reported case has ever tested the PBC balancing duty, bringing further uncertainty to this recourse. In OpenAI's case, one additional hurdle stands in the way of a derivative action: the shareholder best positioned to sue is the controlling non-profit, with its 26% stake, which is also the entity whose financial interests are most aligned with the commercial trajectory the public benefit duty is meant to counterbalance. The result is plaintiffs without a claim in the case of a traditional corporation like Alphabet, or a claim with few plaintiffs for PBCs like OpenAI and Anthropic.

Shareholders may have an additional recourse in the event of a catastrophic failure. Directors must make a good faith effort to implement and monitor oversight mechanisms.[274] After Boeing's safety failures led two aircrafts to crash, courts approved a settlement based on shareholders derivative actions against the board.[275] Though such claims rarely succeed, the directors of AI companies could be held personally liable should they fail to oversee AI systems that lead to major casualties.

2. Investor Protection for Publicly Listed Companies

Federal securities laws oversee publicly listed companies. Companies trading shares on the stock market have to make various disclosures allowing investors to make informed decisions. As part of that process, companies list risk factors.[276] This makes the risk section of 10-K yearly filings of publicly listed

[272] *Id.* at 82, 91.
[273] *Certificate of Incorporation of OpenAI Group PBC*, *supra* note 89, arts. VII-XI; *Anthropic Certificates of Incorporation*, *supra* note 3 Arts. IX-X; Del. Code Ann. tit. 8, § 102(b)(7).
[274] *In re Caremark International Inc. Derivative Litigation*, 698 A.2d 959 (Del. Ch. 1996).
[275] *In re The Boeing Company Derivative Litigation*, 2021 WL 4059934 (Del. Ch. 2021).
[276] SEC, *Investor Bulletin: How to Read a 10-K*, https://www.sec.gov/files/reada10k.pdf.

AI companies a good starting point to understand company practices that may jeopardize broader public interests. Should there be inconsistencies between these statements and actual practices, concerned employees can avail themselves of whistleblower protection to raise unaddressed concerns. The employees of other technology companies have used this route to denounce misrepresentations between internal practices and these public disclosures.[277] For example, ex-Facebook employee Frances Haugen has disclosed multiple ways in which the company misled investors as to its practices regarding content moderation and the effect of its products on teenagers,[278] essentially leveraging disclosure obligations to denounce problematic practices.

However, employees can face retribution for speaking out. Technology workers voicing safety concerns put their equity on the line,[279] a significant portion of their earnings.[280] Some exceptionally principled employees may forego advantages that can amount to nearly a million dollars for top-level engineers, but such financial pressures likely stifle criticism. Some prominent AI companies have walked back provisions in employment contracts causing employees to lose their equity when speaking out,[281] but the residual risk likely chills employees' ability to speak freely.

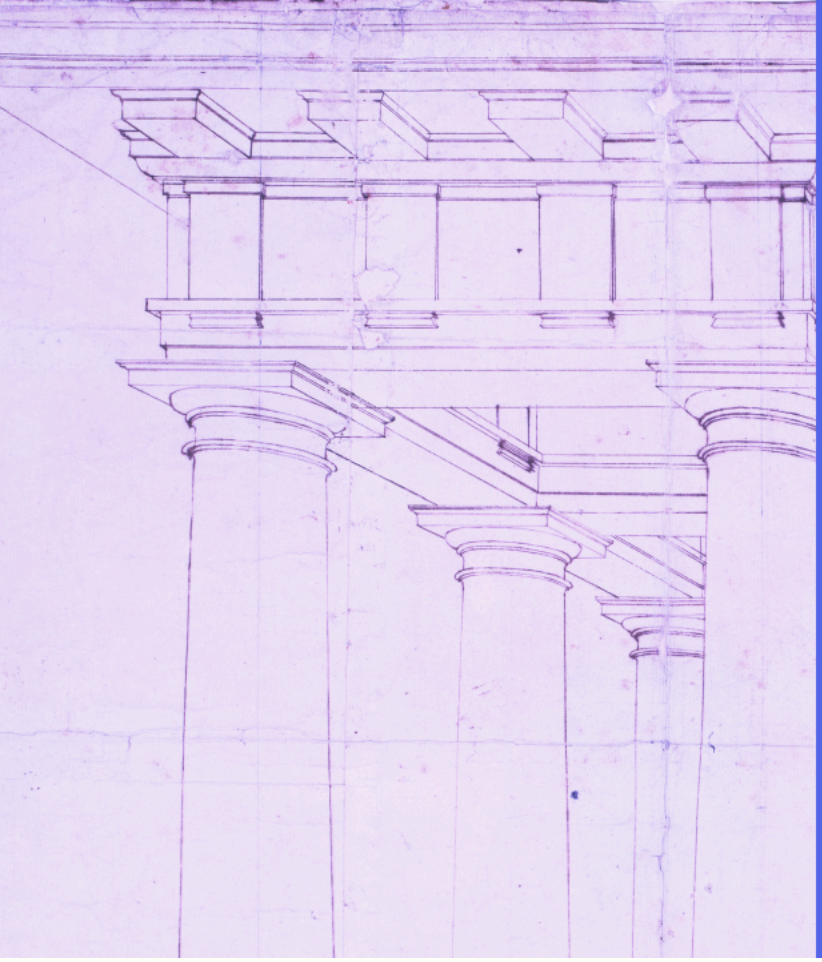
Concretely, Google's disclosure acknowledges that some of its AI activities might result in legal liability, harmful content and other issues, noting that it devotes "significant resources to develop, test, and maintain our products and services, but we may not be able to identify or resolve all AI-related issues, deficiencies, and/or failures before they arise."[282] The Google disclosure is careful to broadly acknowledge AI-related risks, giving little grip to a claim of inconsistency between disclosure and practice. However, it also mentions devoting "significant resources" for development and testing. Should it depart from its scaling policy, this could be a hook to flag an incomplete or misleading disclosure.

---

[277] U.S. Sec. Exch. Comm'n, Whistleblower Protections, https://www.sec.gov/enforcement-litigation/whistleblower-program/whistleblower-protections (last visited Nov. 28, 2025).
[278] Laura Kusisto & Mengqi Sun, *The Facebook Whistleblower, Frances Haugen: Does the Law Protect Her?*, Wall St. J. (Oct. 5, 2021), https://www.wsj.com/finance/regulation/the-facebook-whistleblower-frances-haugen-does-the-law-protect-her-11633453696 [https://perma.cc/6L7T-VXQC]; *The Facebook Files*, Wall St. J. (Oct. 1, 2021), https://www.wsj.com/articles/the-facebook-files-11631713039; April F. Doss, *SEC Complaints: Fallout from the Facebook Files—Part 3*, Just Sec. (Oct. 12, 2021), https://www.justsecurity.org/78553/sec-complaints-fallout-from-the-facebook-files-part-3/.
[279] Kevin Roose, *OpenAI Insiders Warn of a 'Reckless' Race for Dominance*, N.Y. Times (June 4, 2024), https://www.nytimes.com/2024/06/04/technology/openai-culture-whistleblowers.html [https://perma.cc/Y5DA-M7D7]; *see also* Matthias Bastian, *OpenAI's AI Safety Teams Lost at Least Seven Researchers in Recent Months*, Decoder (May 18, 2024), https://the-decoder.com/openais-ai-safety-teams-lost-at-least-seven-researchers-in-recent-months/ [https://perma.cc/9BVZ-CMDB] (citing one OpenAI ex-employee claiming he gave up 85% of his family net worth by foregoing equity and speaking out on safety concerns).
[280] Nguyen, *supra* note 62; Bruce Brumberg, *Amazon, Apple, Google: Restricted Stock Grants Evolve To Retain Top Talent*, Forbes (Apr. 14, 2022), https://www.forbes.com/sites/brucebrumberg/2022/02/08/amazon-apple-google-restricted-stock-grants-evolve-to-retain-top-talent/ [https://perma.cc/WZ7H-49F8]; *OpenAI Software Engineer Salaries,* Levels.fyi (Jan. 27, 2025), https://www.levels.fyi/companies/openai/salaries/software-engineer [https://perma.cc/4BXJ-CYUF] (describing the highest echelon consists in a $379,000 base salary and $960,000 in equity with a four-year vesting schedule).
[281] Shirin Ghaffary & Rachel Metz, *OpenAI Releases Former Staffers From Non-Disparagement Clauses*, Bloomberg (May 24, 2024), https://www.bloomberg.com/news/articles/2024-05-24/openai-releases-former-staffers-from-nondisparagement-clauses; *see also* Bruce C. Judge, *NLRB Decision Invalidates NDAs and Non-Disparagement Clauses*, Whistleblower L. Collaborative (May 2, 2023), https://www.whistleblowerllc.com/nlrb-decision-invalidates-ndas-and-non-disparagement-clauses/ (summarizing recent increased employee protection for speaking out).
[282] *Alphabet Investor Relations - Investors - SEC Filings*, 16, https://abc.xyz/investor/sec-filings/default.aspx (last visited Nov. 19, 2025)2025 10-K filing.

Stock exchange rules add another layer of disclosure requirements. The Nasdaq exchange, where Alphabet stock primarily trades, requires companies to disclose information affecting the value of shares or influencing investors' decisions.[283] Should OpenAI and Anthropic proceed with plans to publicly trade their stock, they would be subject to those requirements as well. Overall, these mandatory disclosures function as information-forcing mechanisms that incentivize companies to minimize practices counter to public interest that may result in reputational harm. However, employees still face significant risks when denouncing inconsistencies between these disclosures and actual practices.

[283] Nasdaq Stock Market LLC Rules, Rule 5250(b)(1) (2025), https://listingcenter.nasdaq.com/rulebook/nasdaq/rules/nasdaq-5200-series (last visited Nov. 19, 2025).



Protecting the Charitable Mission of a Non-Profit

After multiple twists and turns in its corporate structure, the newly-minted OpenAI public benefit corporation remains controlled by a charity, the OpenAI Foundation. Throughout the evolution of the company, the obligation to pursue the charitable purpose embedded in the company's DNA has been a backstop, preventing it from abandoning its mission altogether.

1. OpenAI's Non-Profit Purpose and Ensuing Legal Obligations

OpenAI's founding documents declare that "[t]he specific purpose of this corporation is to provide funding for research, development and distribution of technology related to artificial intelligence. The resulting technology will benefit the public and the corporation will seek to open source technology for the public when applicable. The corporation is not organized for the private gain of any person."[284] The CEO reiterated the crucial role of the non-profit constraint to preserve the company's public interest mission: "At OpenAI, when we wrote our charter, we talked about the scenarios where we would or wouldn't make money. And . . . the things we wouldn't be willing to do no matter how much money they made. And we made this public so the public would hold us accountable to that. And I think that's really important."[285]

---

[284] OpenAI, *Certificate of Incorporation and Bylaws*, *supra* note 40.

[285] Eric Johnson, *Full Q&A: Y Combinator's Sam Altman and Recode's Kara Swisher Discuss Tech Ethics, Addiction and Facebook*, Vox (Dec. 10, 2018), https://www.vox.com/2018/12/10/18134926/sam-altman-kara-swisher-recode-decode-live-mannys-podcast-transcript-facebook-zuckerberg-ethics.

The law of non-profit organizations provides mechanisms to operationalize these commitments. The board of directors and the officers of a charitable organization owe special duties to the non-profit and its donors. Among these obligations is the fiduciary duty of loyalty to put the interest of the non-profit above one's own. A widely cited opinion conveys the high standard of the duty of loyalty as follows:

> "Many forms of conduct permissible in a workaday world for those acting at arm's length, are forbidden to those bound by fiduciary ties. A trustee is held to something stricter than the morals of the market place. Not honesty alone, but the punctilio of an honor the most sensitive, is then the standard of behavior. As to this there has developed a tradition that is unbending and inveterate."[286]

Loyalty requires board members to prioritize the non-profit mission, in good faith and in a manner they reasonably believe to be in the best interests of the charity in light of its purposes.[287] Therefore, the non-profit purpose is the North Star for any decision the board of the Foundation makes. In the new PBC structure, the non-profit board is one degree remote from overseeing the operations of the company, as its role is essentially to elect the board of the PBC (see Figure 7 above). In that sense, the breadth of the duty of loyalty is quite limited. However, the Foundation's board retains oversight of the Safety and Security Committee, and in that capacity could make consequential decisions to halt the release of models or AI systems. Those decisions should comport with the non-profit mission to the exclusion of financial considerations. In the event that they don't, this could be construed as a breach of the duty of loyalty. In California where OpenAI carries most of its activities, the law also creates a fiduciary relationship between a donor and the charity or person soliciting on behalf of that charity.[288] Accepting charitable contributions creates a duty to use the contributions for the charitable purposes for which they were sought. These various obligations binding the Board of the Foundation require it to advance, not stifle, the non-profit mission.

2. Who Can Enforce OpenAI's Non-Profit Mission

State attorneys general play a key role in ensuring that OpenAI remains loyal to its charitable mission. They have broad powers to monitor non-profits converting to for-profit status, including authority to initiate legal proceedings to oppose structural changes that would jeopardize the non-profit's assets and mission.[289] Delaware also specifically bars converting a nonstock corporation "if the charitable status of such charitable nonstock corporation would thereby be lost or impaired."[290]

---

[286] *Meinhard v. Salmon*, 164 N.E. 545, 546 (N.Y. 1928).
[287] A.L.I., Restatement of the Law - Charitable Nonprofit Organizations (2007) § 2.02 "Duty of Loyalty"; Aprill, Loui & Horwitz, *supra* note 71, at 293; Loui, *supra* note 87.
[288] Cal. Bus. & Prof. Code § 17510.8.
[289] Cal. Gov't Code § 12598(a) (2023); Del. Code Ann. tit. 8, § 284 (2025) (authorizing AG to seek charter revocation post facto); Restatement of the Law, Charitable Nonprofit Organizations § 5.01 (Am. Law Inst. 2021).
[290] Del. Code Ann. tit. 8, § 266(j) (2025).

The board members of the charity (formerly OpenAI Inc., now OpenAI Foundation) could also bring a claim on its behalf.[291] But since the charity stands to gain from increased profitability, board members have little incentive to enforce the mission through a derivative action. As for ousted members, Delaware denies them standing to sue on behalf of the charity, even when confronted with the objection that this creates a perverse incentive to oust internal critics as a way to preclude derivative actions.[292] The Restatement of the Law, Charitable Nonprofit Organizations, an authoritative synthesis of charity law which guides but does not bind courts, takes the opposite view: board members removed for trying to address the very harm they seek to challenge should keep their standing.[293] OpenAI's fired directors are a case in point: removed amid a dispute over whether the company was straying from its mission, they lost standing to sue on behalf of the charity.

Donors also have various recourses against the charity. They can invoke a breach of the above-mentioned fiduciary duty to use the collected funds for the intended charitable purpose. Elon Musk's lawsuit against OpenAI detailed below significant hurdles, but they rest on procedural technicalities rather than the substance of the claims.[294] More controversially, donors could try to sue on behalf of the company to enforce the fiduciary duties of the board. In jurisdictions like Delaware, only the attorney general has the ability to do so, but the law is in flux.[295]

Irrespective of the formal standing to sue on behalf of the charity, many actors have sought to prevent changes to the structure of OpenAI that would jeopardize commitments to its mission. They have done so by way of formal court proceedings, but also by using these obligations to channel their demands that OpenAI remain consistent with its mission. The next section outlines how various actors have invoked the spirit and the letter of non-profit law to align the company's action with its mission.

---

[291] A.L.I., *supra* note 287, § 6 "Standing of a Private Party to Bring an Action on Behalf of a Charity: The Derivative Action"; Aprill, Loui & Horwitz, *supra* note 71, at 295 (noting that technically the members of the corporation, not the board members, have standing in Delaware, but that in practice the board members are also the corporation members); *see also* Certificate of Incorporation and Bylaws, *supra* note 40, art. II 1(1) establishing that the board members serve as members of the corporation.

[292] *Gassis v. Corkery*, C.A. No. 8868-VCG, 2014 WL 3565418, at *4–6 (Del. Ch. July 21, 2014) (letter op.), *aff'd*, 116 A.3d 1080 (Del. 2015); *see also Schoon v. Smith*, 953 A.2d 196, 205–08 (Del. 2008) (reserving equitable derivative standing for cases presenting a "complete failure of justice" and declining to extend it to directors).

[293] A.L.I., *supra* note 287, § 6.02(b)(2)(B).

[294] *Musk v. Altman,* No. 4:24-cv-04722 (N.D. Cal. Aug. 5, 2024) (finding by jury that the claims were barred as Musk waited too long to file).

[295] A.L.I., *supra* note 287, § 6.03 Standing of a Private Party to Enforce the Purposes to Which Charitable Assets are Devoted and Administrative Terms Governing Charitable Assets—general comments; *Oberly v. Kirby*, 592 A.2d 445, 464 (Del. 1990) ("Delaware law unambiguously places the burden of protecting the interests of beneficiaries upon the Attorney General."); *Wier v. Howard Hughes Medical Institute*, 404 A.2d 140, 145 (Del. Ch. 1979) ("It is generally recognized that the appropriate person to enforce the provisions of a Delaware charitable trust is the State Attorney General."); Samuel D. Brunson, *Musk, OpenAI and the Internal Affairs Doctrine*, Nonprofit L. Prof Blog (Sep. 10, 2025), https://web.archive.org/web/20250910172001/https://lawprofessors.typepad.com/nonprofit/2024/03/musk-openai-and-the-internal-affairs-doctrine.html (noting that California may be more flexible).

3. Enforcing the Mission in Practice

OpenAI's non-profit mission has allowed a variety of actors to contest every step of the company's commercial pivot. The examples below show that the charitable purpose feeds both formal lawsuits and broader challenges to corporate restructuring away from the original non-profit form.

Investor Lawsuit Contesting the Capped-Profit Joint Venture

When OpenAI pivoted to a capped-profit structure, investor Elon Musk pushed back.[296] Based on donations in the range of $45-50 million,[297] Musk challenged the company's capped-profit model both in the court of public opinion and in actual court. He criticized OpenAI's business model for funding initial R&D with donations, only to slide the resulting IP into for-profit ventures. When the company secured an exclusive licensing deal with Microsoft in return for its $10 billion investment in 2023,[298] Musk alleged that OpenAI, Microsoft and other defendants "have chosen to use GPT-4 not for the benefit of humanity, but as proprietary technology to maximize profits for literally the largest company in the world".[299] Musk narrowed his claims to unjust enrichment and breach of charitable trust shortly before trial. The latter claim rests on the idea that OpenAI owes donors a fiduciary duty to use donations for the intended charitable purposes. The complaint frames the exclusive licensing deal with Microsoft and failure to release details of the GPT-4 model as running afoul of the mission to open source technology for the public benefit.[300] OpenAI's fiduciary duty towards its donors thus played a key role in articulating concerns about mission drift in the capped-profit restructuring.

At the time of finalizing this report, a trial judge dismissed the lawsuit on a technicality, as a jury found that Musk had exceeded the three-year window before filing his complaint.[301] Musk has vowed to appeal and the case is currently assigned a mediator.[302] Irrespective of the lawsuit's outcome, it has put a significant hurdle in the way of OpenAI. The very fact that the company had to expend time and resources on a claim credible enough to make its way to trial, at a critical juncture with an impending IPO and as the company was elbowing with Anthropic catching up to its market position, suggests that legal protections for non-profit organizations put real pressure on business decisions to depart from the stated mission.

---

[296] *See supra* p. 5 (summarizing the change from an initial single non-profit entity towards a joint venture between the original non-profit and newly created for-profit companies).
[297] Harris, *supra* note 49; *OpenAI and Elon Musk*, https://openai.com/index/openai-elon-musk/ (last visited Sep. 24, 2025).
[298] *Musk v. OpenAI,* No. CGC-24-6127476 (Cal. Super. Ct. filed Feb. 29, 2024).
[299] *Id.* ¶ 90.
[300] *Id.* ¶ 134.
[301] Cade Metz & Mike Isaac, *Elon Musk Loses $150 Billion Suit Against OpenAI and Sam Altman*, N. Y. Times (May 18, 2026), https://www.nytimes.com/2026/05/18/technology/elon-musk-lawsuit-openai-sam-altman.html.
[302] Peter Blumberg, *Musk-Altman Judge Enlists Mediator to Resolve AI Titans' Battle*, Bloomberg (July 2, 2026), https://news.bloomberglaw.com/litigation/musk-altman-judge-enlists-mediator-to-resolve-ai-titans-battle.

Broad Pushback Against For-Profit Restructuring

As soon as it was rumored to convert to a for-profit model, OpenAI faced a barrage of criticism.[303] Donors, competitors and civil society coalitions all raised the immutable non-profit mission as a roadblock standing in the way of this change, invoking various aspects of the legal framework pertaining to non-profit organizations. Echoing Musk's contention that diverting donation towards for-profit activities breached the fiduciary duty to use the funds for the stated charitable purposes of open sourcing AI technology for the benefit of humanity,[304] competitor Meta also took issue with OpenAI's about-face. Its complaint to the California Attorney General reads as follows:

> "[I]f OpenAI's new business model is valid, non-profit investors would get the same for-profit upside as those who invest the conventional way in for-profit companies while also benefiting from tax write-offs bestowed by the government and, ultimately, the public. That would distort the market by essentially requiring any startup seeking to remain competitive to adopt the same playbook."[305]

Though it did not cite specific provisions or case law pertaining to the protection of the non-profit purpose, Meta's position is rooted in the principle that a non-profit organization must remain true to its charitable purpose.

Civil society actors also leveraged non-profit rules to oppose OpenAI's for-profit restructuring. An open letter by the Not For Private Gain coalition composed of former OpenAI employees, technology workers, academics and civil society groups warned that removing non-profit control would contravene the legally protected charitable purpose and contravene the fiduciary duties owed to the beneficiaries of the non-profit.[306] Non-profit law concepts thus provided the coalition with the vocabulary and tools to prevent OpenAI from abandoning its non-profit structure and mission altogether.

Legislative intervention also sought to protect the original non-profit purpose of OpenAI. A coalition of researchers, consumer advocates, non-profits, and community leaders called for legal reforms preventing the for-profit conversion of non-profits like OpenAI. The California bill AB 501 would have prohibited non-profits worth over $1 million from converting into for-profit companies.[307] Though the effort did not come to fruition,[308] it shows that the integrity of a non-profit is a potent argument to ensure AI companies structured as such remain true to their mission.

Overall, mobilizing non-profit law to oppose OpenAI's for-profit conversion appears to have had some bearing on the company's decision not to move forward with a full for-profit restructuring. It stated that

---

[303] *See supra* notes 86–88 and accompanying text on the proposed for-profit structure.
[304] Complaint ¶¶ 316–28, *Musk v. Altman*, No. 4:24-cv-04722 (N.D. Cal. Aug. 5, 2024).
[305] Letter from Meta Platforms, Inc., to Rob Bonta, Att'y Gen. of Cal. (Dec. 2024) (on file with author).
[306] Not For Private Gain, *supra* note 88 April 2025 letter.
[307] *AB 501- AMENDED*, https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB501 (last visited Sep. 4, 2025); Peter Windberger, *New Bill AB 501 to Prevent OpenAI's Non-Profit Conversion*, LessWrong (2025), https://www.lesswrong.com/posts/n3fWYBydMi3yGNyLh/new-bill-ab-501-to-prevent-openai-s-non-profit-conversion.
[308] Gary Marcus, *BREAKING: Bill That Would Have Blocked OpenAI's Conversion to a for-Profit Has Mysteriously Been Gutted.*, Marcus on AI (Apr. 6, 2025), https://garymarcus.substack.com/p/breaking-bill-that-would-have-blocked.

"[w]e made the decision for the nonprofit to retain control of OpenAI after hearing from civic leaders and engaging in constructive dialogue with the offices of the Attorney General of Delaware and the Attorney General of California,"[309] thus indicating that non-profit safeguards constrained its ability to shed the charitable form and mission.

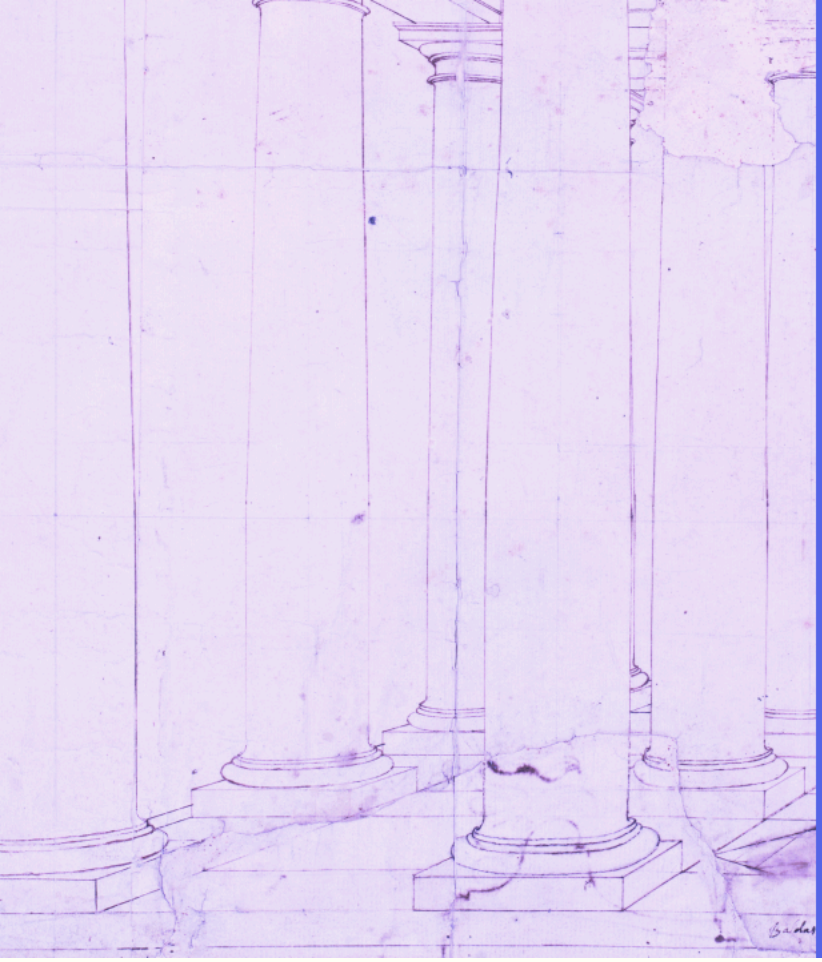
Civil Society and Attorneys General Influence on the PBC Recapitalization Plan

As of November 2025, OpenAI has completed its transition to yet another bespoke model. The non-profit entity retains control over a newly created public benefit corporation, nominating its directors and retaining a right of approval over safety and security matters.[310] In the lead-up to this change, civil society groups have invoked the non-profit purpose to voice concerns over the restructuring as endangering the company's mission. In a letter to the attorneys general of California and Delaware, the Not for Private Gain coalition flagged that the new structure removed the primacy of the charitable purpose, undermining the leadership's fiduciary duty to act in pursuit of the company's mission.[311] A coalition of over 60 organizations has filed a formal petition demanding that the California Attorney General investigate OpenAI's compliance with its charitable mission.[312] The coalition noted that the restructuring would jeopardize the company's legal duty to prioritize its public interest mission, adding that the non-profit should receive adequate compensation for its assets. Both groups channeled non-profit law to align OpenAI with its mission. At their core, they make two claims: the company should remain committed to its charitable mission, and it should use its assets to serve the public good. The attorneys general of Delaware and California secured safety and governance commitments as a condition not to oppose the restructuring, something the section on AG oversight returns to.[313]

Overall, legal safeguards pertaining to non-profit organizations have enabled different stakeholders to contest OpenAI's for-profit conversion, and have secured some protections of the mission in the new PBC structure. However, there are persisting concerns that these measures are insufficient to align the company's behavior with the spirit of its publicly minded mission. As the next section shows, enforcing the mission of a public benefit corporation is a fraught endeavor.

---

[309] *Evolving OpenAI's Structure*, OpenAI, *supra* note 44.
[310] *See supra* note 99 and accompanying text for a discussion of the new structure, the powers of the Foundation's board and the right of approval of its Safety and Security Committee.
[311] Not For Private Gain, *supra* note 88 May 12, 2025 letter. *See* CA MOU, *supra* note 99; Del. Code Ann. tit. 8, § 365(a); *Benefit Corporations: Upholding the Triple Bottom Line*, USC Gould School of Law, https://gould.usc.edu/news/benefit-corporations-upholding-the-triple-bottom-line/ (last visited Nov. 14, 2025) (explaining that Delaware public benefit corporations give directors leeway to balance several factors).
[312] Letter from Fred Blackwell, CEO, San Francisco Foundation et al., to Rob Bonta, Att'y Gen. of Cal. (June 26, 2025), https://docs.google.com/document/d/122d3w4nW9HWLEx4EMOnHEinbezNBsw0t9zdviIK8I-w/edit? (last visited Sep. 23, 2025).
[313] *Supra* note 99.



Enforcing the Mission of a Public Benefit Corporation

OpenAI and Anthropic's public benefit corporations offer several avenues to align their activities with their stated mission. Directors should consider the mission among other factors, while reporting requirements and certifications incentivize mission compliance. Yet in practice, enforcing the mission remains elusive.

1. Limited Director Liability for Prioritizing Profit Over Mission

Whereas traditional for-profit corporations exclusively serve stockholders' interests, public benefit corporations require directors to balance the financial interests of shareholders with the benefits of those materially affected and the specific public interest mission of the company.[314] Directors pursue a triple bottom line: profit, people, planet.[315] But, crucially, they have significant leeway in how they balance these factors. Under the business judgment rule, courts will not second-guess how directors balance the three relevant factors, deferring to them as long as their decisions are informed and disinterested.[316] Directors simply have to clear the low bar of avoiding decisions that "no person of ordinary, sound judgment would approve."[317] Given this significant leeway, PBC directors can still prioritize profits at the expense of the mission,[318] with the notable exception of OpenAI, which must

[314] Del. Code Ann. tit. 8, § 365(a).
[315] *Benefit Corporations: Upholding the Triple Bottom Line*, *supra* note 311.
[316] Del. Code Ann. tit. 8, § 365(b); The Delaware Way, *supra* note 102; Briana Cummings, *Benefit Corporations: How to Enforce a Mandate to Promote the Public Interest*, 112 Colum. L. Rev. 578 (2012) (noting the limited role of shareholder suits given business judgment rule).
[317] Del. Code Ann. tit. 8, § 365(b); *see also* Michael B. Dorff, James Hicks & Steven D. Solomon, *The Future or Fancy? An Empirical Study of Public Benefit Corporations*, 11 Harv. Bus. L. Rev. 113 (2021).
[318] Paul Oudin & Teodora Groza, *The Governance of AI Companies: Reconciling Purpose with Profits* (Sep. 20, 2024) (unpublished manuscript), https://papers.ssrn.com/abstract=4972751.

exclusively consider the mission for safety and security issues, something section 3 below unpacks further.[319]

PBC directors who fail to balance the three factors face minimal consequences. Both companies protect directors from personal liability as much as the law permits,[320] insulating them from punishable bad faith or breach of the duty of loyalty.[321] Further weakening the grip of fiduciary duties is the requirement that shareholders own at least 2% of equity to force the directors of a Delaware PBC to pursue the mission.[322] Should the PBCs go public,[323] that threshold would be lowered to owning $2 million in shares, exposing AI PBCs to more litigation. For the time being, current investors are primarily venture capitalists and large technology companies with few incentives to enforce the mission over profits. Finally, the statute makes clear that directors don't owe a duty to materially affected people or the public writ large.[324] Shareholders, not stakeholders, can hold directors accountable. As a result of all these limitations, few people can demand that directors enforce the mission. As a result, directors get a blank check allowing them to prioritize profit or the mission as they see fit, without fear of legal liability. By some accounts, AI companies have actually flipped the PBC idea on its head: designed to free directors to choose mission over profit, it actually shields directors that prioritize profit over mission.[325]

2. Public Reporting

Public benefit corporations have to report how they promote their stated public benefit mission as well as the interests of those materially affected. That said, the reports are geared towards stockholders, with no statutory obligations to make them public.[326] As part of its deal with the AGs, OpenAI will provide semi-annual reports, but Anthropic's remain confidential.[327] Either way, reporting offers limited post facto accountability, at best incentivizing mission alignment through indirect pressure.

---

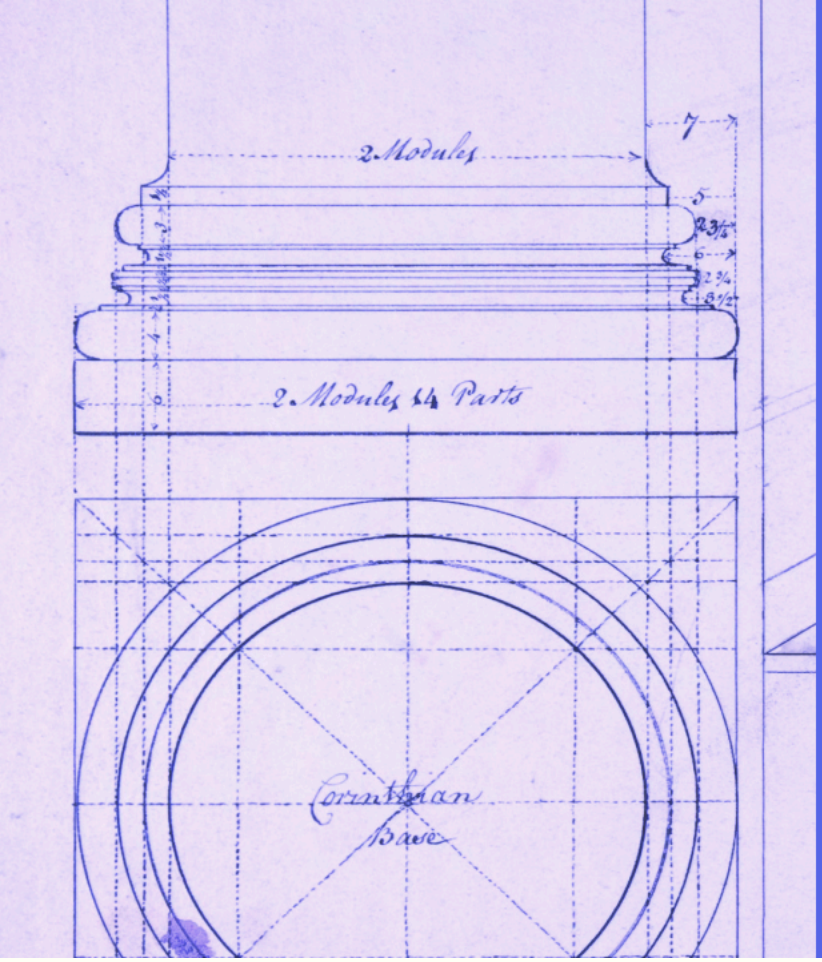
[319] Tyler Whitmer, *OpenAI's Restructuring Makes Safety-Focused Shareholder Litigation Possible*, LASST (Nov. 19, 2025), https://lasst.org/2025/11/19/the-litigation-risk-openai-just-created/ (predicting this will facilitate shareholder actions against the Board for failing to prioritize the mission in safety and security decisions); *see also supra* note 100 and accompanying text for a discussion of the mission-only carve-out.
[320] *See supra* note 273 and accompanying text.
[321] Del. Code Ann. tit. 8, § 365(c); *Anthropic Certificates of Incorporation*, *supra* note 3, Art. IX (A) (Mar. 27, 2025 version); Anne-Marie Beliveau et al., *Delaware Public Benefit Corporations—Recent Developments*, Harv. L. Sch. F. on Corp. Governance (Aug. 31, 2020), https://corpgov.law.harvard.edu/2020/08/31/delaware-public-benefit-corporations-recent-developments/ (noting the change from an opt-in liability limitation to an opt-out regime).
[322] Del. Code Ann. tit. 8, § 367.
[323] George Hammond, *Anthropic Files for Blockbuster Initial Public Offering*, Fin. Times (June 1, 2026), https://www.ft.com/content/4f82f41c-24e7-4323-899a-17a04badd29e?syn-25a6b1a6=1.
[324] Del. Code Ann. tit. 8, § 365(b).
[325] Andrew Muchin Goldblatt, *The Inverted Benefit Corporation: OpenAI and the Limits of Mission Governance* (May 11, 2026) (unpublished manuscript), https://papers.ssrn.com/abstract=6613658 (observing that OpenAI's recapitalization into a PBC introduced profit considerations into a non-profit structure).
[326] Del. Code Ann. tit. 8, § 366.
[327] CA MOU, *supra* note 99 ¶ 17; DE MOU, *supra* note 99 ¶ 15.

3. Attorneys General Oversight

In a public letter citing their ability to block the restructuring, state AGs noted concerns about the role of AI chatbots in the death of children, urging the company to "proactively and transparently ensure AI's safe deployment. Doing so is mandated by OpenAI's charitable mission, and will be required and enforced by our respective offices."[328] Leveraging their power to challenge OpenAI's PBC conversion from a non-profit to a for-profit model, the attorneys general of Delaware and California secured governance and safety commitments in memoranda of understanding of "non-objection" in which they undertake not to take enforcement actions against the company in return for its agreement to a series of conditions.[329]

The agreement requires the OpenAI Foundation to act consistently with the mission, including in their appointment and removal of the PBC Board, which must remain majority independent.[330] Should the company stray from said obligations, the AGs could intervene.[331] The agreement also grants the Safety and Security Committee (a subset of the Non Profit Foundation) a right of approval over safety decisions. This right sits in a contract between the Non Profit Foundation and OpenAI PBC, with the contract itself a condition of the AGs' non-objection.[332] The AGs thus have the option to intervene should the PBC fail to honor the Committee's right of approval, although some have pointed out the vague scope of this right.[333] The non-exhaustive description of the Committee's enforcement power focuses on the release of models or systems, so it's unclear whether it would reach in-house or evaluation environments such as those in which the summer and fall 2026 containment failures occurred.[334] Future research should clarify gaps in the Safety and Security Committee's mandate in light of such incidents.

[328] Letter from Kwame Raoul, Att'y Gen. of Ill. et al., to Bret Taylor, Chair, Board of Directors, OpenAI (Sep. 5, 2025), https://oag.ca.gov/system/files/attachments/press-docs/2025-09-05%20-%20Letter%20from%20DE%20AG%20and%20CA%20AG%20-%20FINAL%20with%20NAAG%20Letter.pdf.
[329] *See supra* notes 289–290 and accompanying text for broad AG powers.
[330] CA MOU, *supra* note 99 ¶¶ 2, 5, 15; DE MOU, *supra* note 99 ¶¶ 1, 4, 13.
[331] Jesse Fried & Idan Reiter, *AI Corporate Governance and Ben & Jerry's Risk* 51 n.221 (Eur. Corp. Governance Inst., Law Working Paper No. 927/2026, 2026), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6697999.
[332] *See supra* note 99 and accompanying text.
[333] Goldblatt, *supra* note 325, at 23, 46.
[334] OpenAI, Hugging Face Incident: Technical Report (Aug. 26, 2026), https://cdn.openai.com/pdf/67869394-cb91-4c12-888c-5cbd85c7814c/OpenAI-Hugging-Face%20Incident-Technical-Report.pdf; METR & Redwood Research, *Brief Independent Investigation of Agents' Behavior, Reasoning and Collaboration in the OpenAI/Hugging Face Hacking Incident* (Aug. 26, 2026), https://metr.org/blog/2026-08-26-openai-hugging-face-incident-investigation/; *An Agent Used DNS to Reach an External Chatbot*, OpenAI (Sep. 25, 2026), https://alignment.openai.com/misalignment-reports/an-agent-used-dns-to-reach-an-external-chatbot/ (documenting an OpenAI agent's September 20, 2026 escape from a sandboxed evaluation environment).



Revoking Corporate Charters

Long dormant, the state's ability to revoke a corporation's ability to operate has recently resurfaced as an option to rein in AI companies. Historically, corporations relied on direct state approval to exist. Governments individually approved the founding documents of corporations, known as charters, allowing them to operate for specific public purposes. In the U.S., corporations enjoyed limited liability in order to undertake major public projects such as the construction of railroads.[335] Corporations carried out public service or infrastructure building with private investment, and the government maintained indirect control through the ability to revoke their charter should they digress from the public goals justifying their existence.[336] The state's power to revoke a corporation's license to operate ensured public control over its activities.[337] In Delaware, where all three companies are incorporated, the Constitution still guarantees the state's ability to revoke the charter of corporations that abuse or misuse their powers,[338] and the relevant statute gives courts jurisdiction to do so for "abuse, misuse or nonuse of its corporate powers, privileges or franchises."[339]

---

[335] For a fuller history of corporations, see Tyler Halloran, *A Brief History of the Corporate Form and Why It Matters*, Fordham J. Corp. & Fin. L. Blog (Nov. 18, 2018), https://news.law.fordham.edu/jcfl/2018/11/18/a-brief-history-of-the-corporate-form-and-why-it-matters/ [https://web.archive.org/web/20210110171139/https://news.law.fordham.edu/jcfl/2018/11/18/a-brief-history-of-the-corporate-form-and-why-it-matters/].

[336] Daniel Hanley, *Pass Corporate Chartering Power to Washington*, Democracy J. (Sep. 8, 2021), https://democracyjournal.org/arguments/pass-corporate-chartering-power-to-washington/.

[337] Thomas Linzey, *Killing Goliath: Defending Our Sovereignty and Environmental Sustainability through Corporate Charter Revocation in Pennsylvania and Delaware*, 6 Dick. J. Env't. L. & Pol'y 31, 36 (1997).

[338] *Delaware Constitution*, https://delcode.delaware.gov/constitution/constitution-10.html (last visited Nov. 21, 2025) ("The General Assembly shall, by general law, provide for the revocation or forfeiture of the charters of all corporations for the abuse, misuse, or non-user of their corporate powers, privileges or franchises. Any proceeding for such revocation or forfeiture, shall be taken by the Attorney-General, as may be provided by law").

[339] Del. Code Ann. tit. 8, § 284(a).

Seldom used today, charter revocation is considered the "nuclear option" reserved for grave misconduct, criminal wrongdoing and statutory violations.[340] Courts put a high bar on revocation, noting in one case that "[c]ontinued serious criminal violations by corporate agents in the course of the discharge of their duties could very well constitute the misuse of a charter," and in another that they look for "sustained course of fraud, immorality or violations of statutory law before deciding that there has been an abuse of charter privileges."[341] That said, some argue that the increased risk of bioweapons is so immoral as to justify revoking the charter of AI companies.[342] This theory remains untested, but opens the door to revoking the charter of any Delaware-incorporated AI company whose actions deviate from the purpose embedded in their corporate form.

OpenAI's charter could also be open to revocation under the theory that its for-profit operations exceed the bounds of the original non-profit corporation. In a case where the organizer operated a non-profit social club for his personal financial benefit, the court observed that "[i]t was not the intention of the Legislature to allow the easy and liberal provisions of the corporation and revenue statutes applicable to corporations organized not for profit to be enjoyed by corporations whose purpose is in fact to engage in business for profit."[343] This theory likely held more purchase during the hybrid capped-profit interregnum than now with the stabilized, well-funded foundation given that the attorneys general have now essentially acquiesced to the new structure.[344] But even if remote, the prospect of a charter revocation may have played a backstop role against a full commercial pivot, as an additional layer of deterrence.

Charter revocation is likely more rhetorical than realistic, especially because a corporation can simply reincorporate in another state.[345] Delaware courts are very unlikely to revoke the charter of a flagship US company as this would erode the state's reputation as a business-friendly jurisdiction—especially at a time when other states are vying to take its place. That said, successful litigation is not the only metric for leverage, as the mere prospect of litigation and even negative public discourse can influence the calculus of how far companies stray from their mission.

Corporate and securities law alone are not a magic bullet, but they can narrow the distance between mission and action. The obligations attaching to non-profit entities have done the most work thus far, effectively preventing OpenAI from completely abandoning its mission. As such, a PBC can foster, but

---

[340] Steven A. Ramirez & Mary Kreiner Ramirez, The Case for the Corporate Death Penalty: Restoring Law and Order on Wall Street (2017), https://muse.jhu.edu/pub/193/monograph/book/57436; Kyle Noonan, *The Case for a Federal Corporate Charter Revocation Penalty*, 80 Geo. Wash. L. Rev. 602 (2012), https://www.gwlr.org/the-case-for-a-federal-corporate-charter-revocation-penalty/; Drew I. Grossman, Note, *Would a Corporate Death Penalty Be Cruel and Unusual Punishment*, 25 Cornell J. L. & Pub. Pol'y 697 (2016), https://scholarship.law.cornell.edu/cjlpp/vol25/iss3/4/.

[341] *Craven v. Fifth Ward Republican Club*, 146 A.2d 400 (Del. Ch. 1958); *Young v. Nat'l Ass'n for Advancement of White People*, 109 A.2d 29, 31 (Del. Ch. 1954); *see also* Edward P. Welch et al., Folk on the Delaware General Corporation Law: A Commentary and Analysis § 284.01, at 10-163 (7th ed. 2021).

[342] Kevin Frazier, *AI, Bioweapons, and Corporate Charters: The Case for Delaware Revoking OpenAI's Charter*, Harv. J. L. & Tech. Dig. Comment. (May 22, 2024), https://jolt.law.harvard.edu/digest/ai-bioweapons-and-corporate-charters-the-case-for-delaware-revoking-openais-charter (last visited Feb. 25, 2025).

[343] Southerland ex rel. Snider v. Decimo Club, Inc., 142 A. 786 (Del. Ch. 1928).

[344] See supra note 98 and accompanying text.

[345] Hanley, *supra* note 336.

not force, a company to align mission and action. Investor protection for publicly listed companies, including conventional for-profit corporations like Alphabet, offers a pathway to hold companies accountable for inconsistencies between their statements and their actions, but employees blowing the whistle stand to lose a significant part of their compensation. As for shareholder resolutions, the dual-class structure of companies like Google cabins them to symbolic gestures. Finally, attorneys general nominally retain the ability to revoke the charter of companies acting inconsistently with their mission or the public interest, but this narrow recourse of last resort is seldom activated.

Conclusion

AI companies take different corporate forms and evolve over time. This report assessed whether function follows form, in the sense of whether certain business structures lend themselves to carrying out publicly minded missions more so than others. The case studies testing how various corporate vehicles walk the talk on their public missions suggest that corporate form has some bearing on how companies approach regulation, but little impact on safety and military policy. When competitive pressures and lucrative contracts are in play, financial interests prevail irrespective of the legal vehicle under which a company operates. In the competitive market that is AI, corporate structure alone cannot align companies with their mission. In fact, those very structures buckle under investor pressures, as is the case with OpenAI gradually abandoning its non-profit ethos to raise more capital.

When it comes to enforcing publicly oriented missions, recourse is limited. With the exception of preventing OpenAI from completely abandoning its original non-profit mission, there are very few instances of material gains. That is not to say efforts to activate corporate law levers are futile; as even the prospect of trouble can, and likely has, constrained the actions of AI companies. Perhaps counterintuitively, time-tested investor protection for publicly traded stock offers more impactful avenues to force disclosures. Should OpenAI and Anthropic proceed with their plan to go public, they, too, will be subject to increased transparency requirements. Accordingly, one important insight and avenue for future advocacy consists in activating disclosure requirements for publicly listed AI companies.

Three ways forward stem from this report. First, current corporate levers are underutilized. For example, OpenAI PBC's board duty to exclusively consider the mission pursuant to its deal with the AGs opens the company to more litigation, especially as the company goes public and shareholders can form coalitions to gather the sufficient stock threshold to proceed with a derivative action on behalf of the PBC. That said, this avenue remains speculative as courts remain extremely reluctant to second-guess directors. Consequently, shareholder activism alone will not align companies with their mission. The limits of current legal vehicles call for new structures that actually have a forcing function on corporate behavior regarding mission compliance. PBCs stop short of that, as they allow but do not require companies to put mission above profits. They are meant to enable profit-driven companies to pursue social goals, not the other way around.

Third and perhaps more importantly, accountability requires robust regulation and public infrastructure. We need laws to move the needle towards publicly beneficial AI. Irrespective of their specific form—and without denying differences in degrees between various corporate vehicles—the self-interest of private companies cannot be meaningfully checked without laws that guarantee companies will factor in the costs of non-compliance. Corporate governance alone cannot ensure companies prioritize their mission over profits. Given the collective action problem that arises from

competition among them (something key players recognized as early as 2019),[346] there is a fundamental mismatch between the incentives of AI companies jockeying for dominance in a capital-intensive market and the broader public interest. As Anthropic's cofounder Jack Clark put it: "I am pretty skeptical of things that relate to corporate governance because I think the incentives of corporations are horrendously warped, including ours. […] It's a real weird thing that this is not a government project."[347] Even OpenAI tried to secure public funding before abandoning its non-profit orientation.[348] Policymakers need to step up with AI regulation that motivates companies to align their actions with the public interest, along with substantial investment in civic-minded public AI infrastructure.

[346] Amanda Askell, Miles Brundage & Gillian Hadfield, *The Role of Cooperation in Responsible AI Development* (July 10, 2019) (unpublished manuscript), http://arxiv.org/abs/1907.04534.
[347] Dylan Matthews, *The $1 Billion Gamble to Ensure AI Doesn't Destroy Humanity*, Vox (July 17, 2023), https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2; *see also* Deep Ganguli et al., *Predictability and Surprise in Large Generative Models*, *in* 2022 ACM Conf. on Fairness Accountability & Transparency 1747 (2022), http://arxiv.org/abs/2202.07785 (noting economic and prestige incentives to move forward even absent adequate assurances).
[348] The Ezra Klein Show, *supra* note 69 (quoting Altman: "A little known fact, we tried to get the public sector to fund us before we went to the capped profit model. There was no interest.").

Annex 1: Google/DeepMind — Structure & Conduct Timeline

Structural Change **Case Study** **Policy/Governance**

Independent startup → absorbed into a for-profit conglomerate

2010	**Private Limited Company** DeepMind founded as an independent startup to "solve intelligence and use it to solve everything else."
2014	**Google Acquisition** Acquired by Google; sale conditioned on an ethics board to preserve the mission. Public assurances surface against military applications
2015	Google dissolves the ethics board after a single meeting
	Alphabet Reorg Alphabet at the apex of the conglomerate; DeepMind operates as Google subset
2018	**Military** Project Maven backlash: after staff protest, Google publishes AI Principles pledging no weapons, overall harm and limiting surveillance
2020 – 22	**Military** Government / defence contracts continue under the AI Principles guardrails
2020 – 24	**Regulation** Urges less restrictive rules in the EU AIA, frames general-purpose regulation as a bad example, opposes external evaluations of high-risk systems
Apr 2023	**Merger** Alphabet merges Google Brain and DeepMind into a single division: Google DeepMind
2024	**Regulation** In California, criticizes SB-1047 for pre-training obligations
Jan 2025	**Restructuring** Gemini teams moved under the single Google DeepMind umbrella; a new ethics council is established, staffed exclusively by Google
Feb 2025	**Military** Lifts the ban on weapons and surveillance applications; drops the reference to "harmful applications."
2025	**Regulation** Supports the federal ban on state AI laws via its trade group; its global policy president criticizes uneven state-by-state protections

Annex 2: OpenAI — Structure & Conduct Timeline

Structural Change **Case Study** **Policy/Governance**

Non-profit → capped-profit hybrid → public benefit corporation

2015	**Non-stock non-profit** Incorporated as OpenAI, Inc., a non-profit counterweight to profit-driven labs. Early stance broadly disfavors military applications
2018	Non-binding Charter commits to a "fiduciary duty to humanity."
2019	**Capped-profit + LLCs** Pivot to a "capped-profit" structure to raise capital; Microsoft partnership begins.
2019 – 24	**Military** Prohibits military and warfare applications and weapons development
2023	Microsoft invests $10B+ for 49% of profits; non-profit keeps formal control. Board fires Altman, then reinstates him within days
2021 – 23	**Regulation** Grow increasingly critical of the EU AIA; threatens to withdraw ChatGPT from EU
2024	**Military** Drops the blanket "military" ban, pursues defence contracts
2024	**Regulation** Opposes SB-1047 and state regulation
2024	Tender offer at a $157B valuation lets employees sell profit-participation units
2025	**Regulation** Pushes for federal preemption; and advocates for a non-binding federal framework
2025	**Scaling/Safety** Preparedness framework commits to pre-deployment safeguards
Oct 2025	**Public Benefit Corp** Restructured as a Delaware PBC; non-profit retains control and owns 26% of stock; valuation ~$300B
Feb 2026	**Military** Offers product to Pentagon for all lawful uses hours after Anthropic's refusal

Annex 3: Anthropic — Structure & Conduct Timeline

Structural Change **Case Study** **Policy/Governance**

PBC with a Long-Term Benefit Trust

Jan 2021	**Delaware PBC** Founded by former OpenAI staff (incl. Dario Amodei) over safety concerns; incorporated as a Delaware public benefit corporation from the outset
2021 – 23	**Regulation** Public commentary supportive of risk-based regulation; engages the EU on high-safety safety goals
2023	**+ Long-Term Benefit Trust** Disinterested trustees take a long view on safety, but profit-driven shareholders retain the ability to roll back the trust
Sep 2023	**Military** First archived Acceptable Use Policy bans weapons development and "battlefield management applications", but never military use as such
Jun 2024	**Regulation** Files a "support-if-amended" letter on SB-1047: opposes prescriptive pre-harm mandates, proposes outcome-based deterrence
Jun 2024	**Military** Usage Policy adds protection of military bases; exceptions allow carve-outs for government customers
Aug 2024	**Regulation** Cautious support SB-1047 after amendments, as benefits likely outweigh its costs
Nov 2024	**Military** Partners with Palantir and AWS to bring Claude to US intelligence and defense agencies in classified environments
May 2025	**Scaling/Safety** Responsible Scaling Policy v2.2 adopts precautionary approach; a high-water mark for voluntary safety measures
Jun 2025	**Regulation** Opposes the proposed 10-year federal moratorium on state AI regulation as "too blunt an instrument", proposing a federal transparency standard instead
Feb 2026	**Scaling/Safety** Walks back safety commitments, citing competitive pressure
Feb 2026	**Military** Conflict with Pentagon over autonomous weapons and mass surveillance

Credits

Thanks to Aashna Agarwal, Sacha Alanoca, Kate Brennan, Jesse Fried, Shira Gur-Arieh, Teodora Groza, Heidy Khlaaf, Martha Minow, Sarah Myers West, Robert Yalden and Tom Zick for comments and feedback, along with Camille Chichet, Oliver Kulik, Nalanda Sharadjaya and Mariana Ramirez Cabal for research and editorial assistance.

The author gratefully acknowledges financial support from the Harvard Law School Summer Public Interest Funding program.

Design by Partner & Partners.