Aligned Structuring of AI Startups

Gad Weiss*

In late 2024, OpenAI announced plans to overhaul its signature "capped-profit" organizational structure, prompting sharp public criticism and calls on the Attorneys General of California and Delaware to intervene. The proposed changes would have affected the governance features that had long set OpenAI apart from conventional Silicon Valley startups: a nonprofit parent with undisputed control over its AI development business, caps on investor and employee returns, and a legal framework requiring managers to prioritize safety over profits. Alongside its rival Anthropic, OpenAI had been one of the most prominent early adopters of what this paper calls Aligned Structuring. These are business structuring models that allocate the startup's control rights and cash flows among its owners and stakeholders in a way that aims to ensure that public interests are given priority over profits when deciding how new AI technologies are developed and deployed.

This paper asks how regulators and policymakers should think about Aligned Structuring and so-called "aligned" AI startups as a matter of public policy. Should other high-impact AI startups be encouraged, or even required, to adopt similar structures? When, if ever, should public authorities intervene to prevent such structures from being rolled back or weakened, as critics urged in OpenAI's case? A close examination of the theory behind Aligned Structuring and the realities of startup finance suggests that its promise as a safeguard for the public interest should be met with skepticism.

Mission-oriented outsider decision-makers often lack the leverage or expertise to counterbalance the startup's entrepreneurs or investors, even

---

* Wagner Fellow, NYU School of Law and NYU Pollack Center for Law & Business. This project is based on research conducted towards a J.S.D. degree at Columbia Law School. For invaluable conversations and comments, I thank Zohar Goshen, Karen Brenner, Danny Cuatrecasas, Stephen Choi, Nizan Geslevich Packin, Adi Marcovich Gross, Joshua Mitts, Paul Oudin, Elizabeth Pollman, Ed Rock, David Schizer, David Yermack, Matt Wansley, and Paul Weitzel. This paper has benefitted greatly from the comments of participants in the 2024 Law & Entrepreneurship Association Annual Meeting; the 2024 Law and Finance of Private Equity and Venture Capital; the 2024 National Business Law Scholars Conference; the Spring 2024 Global Technology Law & Policy Seminar; the Inaugural Roundtable on the Law of AI Safety; the WINIR Young Scholars Online Workshop on AI and Institutions; the Sh[AI]rk Tank Online Workshop; and seminars and colloquia held at Columbia Law School, NYU School of Law, NYU Stern School of Business, and Cardozo School of Law. My thanks to the editors of the *Arizona State Law Journal* for their exceptional editorial work.

when they hold formal control. Profit caps are frequently ineffective or easy to circumvent, especially when equity holders are motivated by strategic rather than purely financial goals or can benefit through indirect channels. Many of the touted advantages of alternative business entities, such as empowering managers to prioritize the public interest or enabling stakeholders to hold them accountable, are already largely available under traditional corporate structures. At the same time, the nonconformity of Aligned Structuring may deter the capital and talent needed for responsible AI innovation, to society's detriment. Policymakers aiming to mitigate AI risks must therefore look beyond the formal features of Aligned Structuring and focus instead on the fact-specific subtleties of how power is actually allocated before deciding whether a given model can effectively protect the public interest.

INTRODUCTION

In April 2025, a public letter titled *Not for Private Gain* was submitted to the Attorneys General of Delaware and California.[1] Signed by leading AI researchers and thought leaders in law and finance (including no less than three Nobel laureates), it urged both Attorneys General to block OpenAI's then-rumored plan to overhaul its signature "capped profit" business structure.[2] The proposed changes would have, among other things, transferred control of OpenAI's AI development business from its nonprofit parent to its for-profit arm, and, reportedly, removed the caps on investor returns.[3] These changes, critics warned, "would eliminate most, if not all, of the governance safeguards that OpenAI has insisted are important to its charitable purpose"[4] —that is, "to ensure that artificial general intelligence . . . benefits all of humanity."[5]

OpenAI was not alone in experimenting with similar structural safeguards against profit-driven and socially harmful AI development. Alongside its chief rival, Anthropic,[6] the two startups pioneered the concept this paper calls *Aligned Structuring*—a deliberate allocation of control and economic rights among a startup's entrepreneurs, investors, and stakeholders, aiming to align its business decision-making with the public interest rather than profit-making. While most Silicon Valley startups are believed to optimize their business structures to attract capital and talent on the best possible terms,[7] these models' main goal is to isolate the startup's decision-makers from shareholder-value-maximizing pressures and ensure that stakeholder and public interests play a meaningful role in shaping the development and deployment of advanced AI technologies.[8]

---

1. Page Hedley et al., *April 17 Letter*, *Not for Private Gain* (Apr. 17, 2025) [hereinafter *Not for Private Gain First Letter*], https://notforprivategain.org [https://perma.cc/CR6S-8U9J].

2. *Id.* ("We request that you stop the restructuring and protect the governance safeguards . . . .").

3. *Id.* § II.A.

4. *Id.* § II.

5. *OpenAI Charter*, OPENAI, https://openai.com/charter [https://perma.cc/4WWB-YLDU].

6. *The Long-Term Benefit Trust*, ANTHROPIC (Sept. 19, 2023) [hereinafter: *LTBT Primer*], https://www.anthropic.com/news/the-long-term-benefit-trust [https://perma.cc/LQV9-T26J]; John Morley et al., *Anthropic Long-Term Benefit Trust*, HARV. L. SCH. F. CORP. GOVERNANCE (Oct. 28, 2023), https://corpgov.law.harvard.edu/2023/10/28/anthropic-long-term-benefit-trust [https://perma.cc/WF37-HL3D].

7. *See* discussion *infra* Part II.

8. OPENAI, *supra* note 5 ("Our primary fiduciary duty is to humanity. We anticipate needing to marshal substantial resources to fulfill our mission, but will always diligently act to minimize conflicts of interest among our employees and stakeholders that could compromise broad benefit."); *LTBT Primer*, *supra* note 6 ("We set out to design a structure that would supply

This paper asks whether Aligned Structuring can and should play a role in AI regulation as a means of protecting society from the risks posed by startups developing advanced AI. Are these models likely to serve as meaningful protectors of society's interests? Should they be treated as a benchmark that all high-impact AI startups ought to meet? Should regulators and policymakers, as the *Not for Private Gain* letter urged, treat rollbacks of such structures as cause for alarm or intervention? To do so, the analysis below will question the underlying logic of Aligned Structuring as a concept: whether its core design features are, in principle, capable of realigning startup behavior with the public interest, and whether they can do so at scale without creating greater harms.

For the adoption of Aligned Structuring models to be desirable as a matter of policy, two conditions must be met. First, these models must work: they must meaningfully shape startup decisions, encouraging attention to safety and ethics issues that would otherwise be neglected or discounted.[9] Second, the model must not destroy the value it aims to protect. If Aligned Structuring weakens AI startups' ability to raise capital and attract the talent they need to bring transformative technologies to the market, the societal cost of the drag on innovation may outweigh any governance benefits.[10] This paper argues that the most visible versions of Aligned Structuring to date, despite their ambition, struggle to meet these tests. Their mechanisms are fragile, their applicability narrow, and their costs too high for most AI startups to bear.

Aligned Structuring models make structural departures from standard startup practice in three distinct areas: the allocation of control rights, the allocation of economic rights, and choice of entity. First, startups are typically controlled by their entrepreneurs, their investors, or both.[11] This allocation of control is crucial for allowing startup shareholders with different risk appetites, investment horizons, and skills to trust each other and collaborate effectively.[12] OpenAI and Anthropic, conversely, have decided to shift a significant extent of control rights away from their entrepreneurs and investors to financially disinterested outsiders who are committed to promoting the public interest (here, "*Alignment Stewards*"). OpenAI's for-profit arm was created as a wholly controlled subsidiary of its nonprofit

---

our directors with the requisite accountability and incentives to appropriately balance the financial interests of our stockholders and our public benefit purpose . . . .").

9. *See* discussion *infra* Part III.

10. *See* discussion *infra* Part III.

11. BRAD FELD & JASON MENDELSON, VENTURE DEALS 77–81 (4th ed. 2019); Elizabeth Pollman, *Startup Governance*, 168 U. PA. L. REV. 155, 181 (2019); D. Gordon Smith, *The Exit Structure of Venture Capital*, 53 UCLA L. REV. 315, 326–27 (2005).

12. *See* discussion *infra* Section II.A.

parent—a research lab on a mission to "advance digital intelligence in the way that is most likely to benefit humanity as a whole."[13] Anthropic granted extensive control to the "Long Term Benefit Trust," an external decision-making body controlled by independent trustees and dedicated to developing AI "for the long-term benefit of humanity."[14]

This transfer of control rests on a fragile premise: that Alignment Stewards will exercise their formal powers in ways more consistent with the public interest than insiders would. Two vulnerabilities undermine that hope. First, the allocation of the startup's formal control rights, such as board seats or veto rights, is only one aspect of the startup's control structure. In practice, informal levers, such as control over the financial capital, computing power, or knowledge assets that the startup needs, can give startups' profit-seeking investors,[15] Big Tech strategic partners,[16] or entrepreneurs[17] significant influence over the startup and its Alignment Stewards' decision-making. As this paper will illustrate, even Alignment Stewards whose interests are perfectly aligned with society's might succumb to pressures from profit-seeking stakeholders if the alternative is watching the startup fail, to society's detriment.[18]

Second, Alignment Stewards' commitment to safety and ethics does not necessarily make them better decision-makers overall, from a public interest perspective, compared to the startup's insiders.[19] Unsafe or unethical conduct

---

13. *Introducing OpenAI*, OPENAI (Dec. 11, 2015), https://openai.com/index/introducing-openai [https://perma.cc/ABR4-C5JZ]; *Our Structure,* OPENAI [hereinafter *Capped Profit Structure Primer*], https://archive.ph/QY5I4.

14. *LTBT Primer*, *supra* note 6.

15. Ronald J. Gilson, *Engineering a Venture Capital Market: Lessons from the American Experience*, 55 STAN. L. REV. 1067, 1074 (2003); PAUL GOMPERS & JOSH LERNER, THE VENTURE CAPITAL CYCLE 171–200 (2d ed. 2006); William A. Sahlman, *The Structure and Governance of Venture-Capital Organizations*, 27 J. FIN. ECON. 473, 506 (1990).

16. Amba Kak et al., *Make No Mistake—AI Is Owned by Big Tech*, MIT TECH. REV. (Dec. 5, 2023), https://www.technologyreview.com/2023/12/05/1084393/make-no-mistake-ai-is-owned-by-big-tech [https://perma.cc/Z7KY-CS6X]; Madhumita Murgia, *Big Tech Companies Use Cloud Computing Arms to Pursue Alliances with AI Groups*, FINANCIAL TIMES (Feb. 5, 2023), https://www.ft.com/content/5b17d011-8e0b-4ba1-bdca-4fbfdba10363 [https://perma.cc/U7J3-ZFS2]; Georg Riekeles & Max von Thun, *AI Won't Be Safe Until We Rein in Big Tech*, EUR. POL'Y CTR. (Nov. 22, 2023), https://www.epc.eu/publication/AI-wont-be-safe-until-we-rein-in-Big-Tech-55e63c [https://perma.cc/KL37-VV4N]. For information about foundation models, see *infra* note 79 and accompanying text.

17. *See infra* notes 223–24 and accompanying text; *see generally* Assaf Hamdani & Kobi Kastiel, *Superstar CEOs and Corporate Law*, 100 WASH. U. L. REV. 1353, 1366–96 (2023).

18. *See* discussion *infra* Section III.A.1.

19. Uri Y. Hacohen, *Policy Implications of User-Generated Data Network Effects*, 33 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 340, 384–85 (2023); Roberto Tallarita, *AI Is Testing*

may indeed result from startup managers prioritizing other considerations over safety and ethics (a conflict problem),[20] but it can also stem from their genuine failures to identify and address safety or ethics issues effectively (a competence problem).[21] Alignment Steward directors may be less conflicted than their profit-seeking colleagues, but their inherent information disadvantage compared to some insider directors, and product-specific facts pertaining to the nature of AI risks at hand, might also make them less competent—particularly when compared to entrepreneur-directors.[22]

The second Aligned Structuring feature analyzed is profit-capping: that is, placing caps on the ability of startup owners to benefit financially from the startup's success. Traditionally, allowing entrepreneurs, investors, and employees the opportunity to earn exceedingly high profits in the event of a successful exit is seen as vital for the startup to be able to attract capital and compete with Big Tech firms over top talent.[23] OpenAI's capped profit structure, however, has placed a ceiling on the financial upside of investors and employees.[24] One apparent goal of the profit-capping mechanism was to reduce the likelihood that its equity holders will prioritize financial performance over the public interest.[25]

Capping profits, however, is doubtful as a means for curbing owners' profit-seeking motivations[26] for two main reasons. First, for profit-capping to affect investor behavior, the investor must be significantly motivated by financial profits. However, when AI startups are concerned, this is often not the case. Unlike traditional startup investors, a critical group of AI investors—Big Tech firms and their corporate venture capital ("CVC")

---

the Limits of Corporate Governance, HARV. BUS. REV. (Dec. 5, 2023), https://hbr.org/2023/12/ai-is-testing-the-limits-of-corporate-governance [https://perma.cc/N8YM-TTC3].

20. *See generally* Zohar Goshen & Richard Squire, *Principal Costs: A New Theory for Corporate Law and Governance*, 117 COLUM. L. REV. 767, 790–95 (2017) (discussing conflict costs considerations in corporate governance).

21. *See generally* Goshen & Squire, *supra* note 20, at 785–90 (discussing competence costs considerations in corporate governance).

22. Darian M. Ibrahim, *Debt as Venture Capital*, 2010 U. ILL. L. REV. 1169, 1192 (2010); Elizabeth Pollman, *Team Production Theory and Private Company Boards,* 38 SEATTLE U. L. REV. 619, 628 (2015). For a similar discussion focused on the governance structures of publicly traded firms, see Zohar Goshen & Assaf Hamdani, *Corporate Control and Idiosyncratic Vision*, 125 YALE L.J. 560 (2016).

23. *See* discussion *infra* Section II.B.

24. *OpenAI LP*, OPENAI (Mar. 11, 2019) [hereinafter *Capped Profit Announcement*], https://openai.com/index/openai-lp [https://perma.cc/A8NG-8ASE]; *Capped Profit Structure Primer*, *supra* note 13.

25. *Capped Profit Announcement, supra* note 24.

26. *See supra* note 16 and accompanying text.

arms[27]—are typically more interested in achieving strategic investment goals, such as early access to new technologies and talent, cultivating new markets for their products and services, or collecting information on potential future collaborations before making a more significant investment commitment.[28] Other equity holders may be motivated by financial returns but use means unrelated to their capped economic rights to obtain them—notably, commercial relationships between the startup and their related parties.[29]Additionally, capping shareholders' upside potential is likely to make them more averse to financial risks generally.[30] With less to gain from fluctuations in the startup's value, profit-capped holders are less likely to support *any strategies* that sacrifice present-day value for potential future gains, including those based on applying stricter safety and ethics standards for their technologies or otherwise protecting the public interest.[31]

The third and final Aligned Structuring feature discussed is the use of alternative business entities. Startups typically organize as Delaware corporations—a decision driven by tax considerations, investors' expectations, and lower transaction costs.[32] However, due to concerns that directors' fiduciary duty to maximize enterprise value would preclude a strong-form commitment to public interest,[33] Anthropic and OpenAI have decided to use business entities with more flexible or inclusive fiduciary

---

27. CB INSIGHTS, *Generative AI Bible: The Ultimate Guide to GenAI Disruption* 23, 52 (2023) [hereinafter *Generative AI Bible*]; PITCHBOOK, GENERATIVE AI VC TRENDS: SEGMENT AND CATEGORY ANALYSIS 8–9 (2023) [hereinafter PITCHBOOK GENERATIVE AI REPORT]. For a discussion of CVCs and their investment theses, see Anat Alon-Beck, *Alternative Venture Capital: The New Unicorn Investors*, 87 TENN. L. REV. 983, 1017–22 (2020); Henry Chesbrough, *Making Sense of Corporate Venture Capital*, HARV. BUS. REV. (Mar. 2002), https://hbr.org/2002/03/making-sense-of-corporate-venture-capital [https://perma.cc/L38X-ULND]; Jennifer S. Fan, *Catching Disruption: Regulating Corporate Venture Capital*, 2018 COLUM. BUS. L. REV. 341, 347–58 (2018).

28. Ilya A. Strebulaev & Amanda Ying Wang, *Organizational Structure and Decision-Making in Corporate Venture Capital* 23 (Stan. Bus. Sch. Working Paper No. 4008, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3963514 [https://perma.cc/3LP3-RWJK].

29. *See generally* Chesbrough, *supra* note 27.

30. The risk-averse tendencies of fixed claimant stakeholders (e.g., debtors) is one of the classic justifications for the shareholder primacy principle in corporate law. *See, e.g.*, Frank H. Easterbrook & Daniel R. Fischel, *Voting in Corporate Law*, 26 J.L. & ECON. 395, 403–04 (1983); Julian Valesco, *Shareholder Ownership and Primacy*, 2010 U. ILL. L. REV. 897, 912–16 (2010).

31. *See* discussion *infra* Section III.B.2.

32. Joseph Bankman, *The Structure of Silicon Valley Start-Ups*, 41 UCLA L. REV. 1737, 1739–40 (1994); Susan C. Morse, *Startup Ltd.: Tax Planning and Initial Incorporation Location*, 14 FLA. TAX REV. 319, 329–33 (2013); Gregg Polsky, *Explaining Choice-of-Entity Decisions by Silicon Valley Start-Ups*, 70 HASTINGS L.J. 409, 411 (2019).

33. *See LTBT Primer*, *supra* note 6; *Capped Profit Structure Primer*, *supra* note 13; *infra* notes 285–89 and accompanying text.

duties. Anthropic was formed as a public benefit corporation ("PBC")[34]—a specialized corporation where directors must balance between shareholders' financial interests and society's broader interests.[35] OpenAI's for-profit arm was formed, until recently, as a Limited Liability Company ("LLC"),[36] in which managers' fiduciary duties can be easily modified or even eliminated.[37]

One of the more visible uses of nonstandard legal entities is to protect directors who prioritize AI safety and ethics from challenges brought by profit-focused shareholders.[38] That concern, however, is largely overstated. Even under a strict shareholder-primacy perception of Delaware's corporate law, directors already enjoy broad discretion to take such considerations into account if it can be plausibly justified as maximizing shareholder value in the long term.[39] When disputes do arise, courts rarely second-guess directors' discretion in this respect,[40] and startups can even further shield directors from such challenges through court-tested private ordering tools such as covenants not to sue.[41] Most importantly, the risk of fiduciary litigation in venture-backed startups is remote due to the cultural norms of Silicon Valley, where reputational concerns strongly discourage adversarial behavior.[42] When the opposite concern arises—meaning, ensuring that safety-minded shareholders can hold directors accountable and enforce a startup's social mission—private ordering solutions within traditional corporate law are also often more than sufficient.[43] Startup charters and shareholder agreements can grant veto rights or other oversight mechanisms to shareholders without requiring a shift to an LLC or PBC form.[44] Using special business entities is, therefore, neither

---

34. *LTBT Primer*, *supra* note 6.

35. *See infra* notes 295–96 and accompanying text.

36. *Capped Profit Structure Primer*, *supra* note 13; OpenAI's for-profit arm was recently restructured as a PBC; see discussion *infra* notes 110–11 and accompanying text.

37. *See infra* note 291 and accompanying text.

38. *See Capped Profit Structure Primer*, *supra* note 13.

39. Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 182 (Del. 1986); Mills Acquisition Co. v. Macmillan, Inc., 559 A.2d 1261, 1282 n. 29 (Del. 1989).

40. *See infra* notes 304–07 and accompanying text; *see generally* DAVID A. DREXLER ET AL., DELAWARE CORPORATION LAW & PRACTICE § 15.03 (2023).

41. New Enter. Assocs. 14, L.P. v. Rich, 295 A.3d 520, 539–93 (Del. Ch. 2023); Transcript of *In re* Altor Bioscience Corp., C.A. No. 2017-0466-JRS (Del. Ch. May 15, 2019) (discussed in *New Enter. Assocs.*, 295 A.3d 587–89); *see infra* notes 309–15 and accompanying text.

42. *See generally* Brian J. Broughman & Matthew Wansley, *Risk-Seeking Governance*, 76 VAND. L. REV. 1357–58 (2023); Jesse M. Fried & Mira Ganor, *Agency Costs of Venture Capitalist Control in Startups*, 81 N.Y.U. L. REV. 967, 999–1001 (2006).

43. *See* discussion *infra* Section III.C.2.

44. DEL. CODE ANN. tit. 8, § 122(18). For a discussion of the use of such veto rights in startups' governance structures to protect venture capital ("VC") investors' interests, see, for example, Smith, *supra* note 11, at 346–47.

necessary to protect safety-conscious directors nor rein in profit-seeking directors.[45]

While Aligned Structuring has not been shown to reliably improve a startup's commitment to the public interest, it imposes real and potentially significant costs. These costs arise because Aligned Structuring departs from the conventional governance and financing frameworks that startups rely on to minimize agency costs and bridge information asymmetries in order to attract capital and talent on favorable terms.[46] Reallocating control rights and economic incentives to prioritize societal interests instead disrupts the very architecture that enables startups to develop new technologies in the first place—*any kind* of technology, whether its impact is net-negative or net-positive from society's standpoint.[47] For example, inserting Alignment Stewards into decision-making or capping investor returns may make the startup less attractive to venture capitalist ("VC") investors, who use these rights to minimize entrepreneur-side opportunism[48] or ensure that entrepreneurs are sufficiently incentivized to stay engaged.[49] At the same time, nonstandard legal forms like LLCs or PBCs, even when technically workable, can trigger tax compliance concerns or friction with VC investors and give a costly competitive edge to other startups competing over those VCs' limited attention span.[50] These burdens may be tolerable for a minority of elite startups with exceptional bargaining power, but for the median AI startup, they risk undermining the innovation pipeline itself.[51]

The bottom line is that Aligned Structuring is not an inherently flawed idea, but it is highly fragile. Its potential benefits depend on specific, often rare conditions, while its costs to startups and to the broader public can be substantial. For entrepreneurs and investors, these structures may undermine incentives, financing prospects, and operational clarity. For society, they risk slowing the development of AI technologies across the board, not just those with unreasonable risk profiles. As a result, there is no reliable basis to assume that an "Aligned" startup is, all things considered, more attuned to

---

45. *See* discussion *infra* Section III.C.2.

46. *See* discussion *infra* Part II; *see generally* Robert P. Bartlett, III, *Venture Capital, Agency Costs, and the False Dichotomy of the Corporation*, 54 UCLA L. REV. 37 (2006); Robert P. Bartlett, III, *Shareholder Wealth Maximization as Means to an End*, 38 SEATTLE U. L. REV. 255, 265–66 (2015); Fried & Ganor, *supra* note 42; Smith, *supra* note 11, at 323–24.

47. *See* discussion *infra* Part D.

48. *See infra* notes 346–51 and accompanying text.

49. *See infra* note 351 and accompanying text.

50. *See* Gregg Polsky, *Explaining Choice-of-Entity Decisions by Silicon Valley Start-Ups*, 70 HASTINGS L.J. 409, 442 (2019); Gad Weiss, *The Venture Corporation*, 62 AM. BUS. L.J. 45, 63 (2025); *infra* notes 355–63 and accompanying text.

51. *See infra* notes 366–72 and accompanying text.

the public interest than one that is not, and certainly no justification for treating it as a governance silver bullet or a benchmark for public policy.

Exploring the potential and limits of Aligned Structuring to protect society from AI-related risks is a timely and worthwhile endeavor. Aligned Structuring continues to evolve within its use by its early adopters—OpenAI itself has eventually reformed its Aligned Structuring model in late 2025.[52] The governance issues OpenAI and Anthropic addressed with these features are also not unique to them. Other high-profile, socially aware AI startups have shown interest in harnessing their governance and capital structures to achieve similar purposes. Notably, xAI, Elon Musk's AI venture, was formed as a Nevada Benefit Corporation rather than a standard corporation.[53] SSI—an intriguing AI venture recently formed by former OpenAI scientist Ilya Sutskever—announced that its "business model means safety, security, and progress are all insulated from short-term commercial pressures,"[54] possibly suggesting the use of similar governance or capital structures. Given the involvement of legal scholars in designing current Aligned Structuring models,[55] expanding the scholarly discussion on how these models work and when they might fail is likely to influence their design and use cases in practice.

OpenAI and Anthropic's structures have attracted business law scholars' interest[56] and even prompted some research works dedicated to understanding them.[57] This paper makes two main contributions to this developing body of literature. First, rather than taking the case study approach used by other

---

52. *See infra* notes 103–10 and accompanying text.

53. Becky Peterson, *Musk's xAI Incorporates as Benefit Corporation With 'Positive Impact' Goal*, THE INFORMATION (Dec. 26, 2023), https://www.theinformation.com/articles/musks-xai-incorporates-as-benefit-corporation-with-positive-impact-goal.

54. SAFE SUPERINTELLIGENCE INC., https://ssi.inc/ (last visited Feb. 5, 2026).

55. Anthropic acknowledged Harvard Law School's Noah Feldman and Yale Law School's John Morley for assisting in designing its model. *See infra* note 121.

56. *See, e.g.*, Alexandra Andhov, *OpenAI's Transformation: From a Non-profit to a 157 Billion Valuation*, 46 BUS. L. REV. 2 (London, 2025); Matteo Gatti, *What Corporate Governance for AI?*, JOTWELL (June 13, 2024), https://corp.jotwell.com/what-corporate-governance-for-ai/ [https://perma.cc/ZDL9-MUBX] (reviewing Tallarita, *supra* note 19); Darryll K. Jones, *Charitalism and Mixed Markets in Federal Tax Exemption: A Case Study Using OpenAI's Joint Venture with Microsoft*, 52 CAP. U.L. REV. 573 (2025); Tallarita, *supra* note 19.

57. *See, e.g.*, Andhov, *supra* note 56; Tallarita, *supra* note 19; Paul Oudin & Teodora Groza, *The Governance of AI Companies: Reconciling Purpose with Profits*, SSRN (Nov. 24, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4972751 [https://perma.cc/ZHH3-YZAK]; Frederik Hovmark Pedersen, *Ownership at OpenAI from The Perspective of Enterprise Foundation Governance*, SSRN (Apr. 15, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4795279 [https://perma.cc/M7UA-FRNZ]; Lawrence J. Trautman & Larry D. Foster, II, *Sam Altman, OpenAI, and The Importance of Corporate Governance*, SSRN (Feb. 8, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4679613.

researchers of Aligned Structuring models, focusing on supposed failures in these models as applied by OpenAI or Anthropic in particular,[58] this paper focuses on Aligned Structuring *as a concept*. Doing so, it aims to help structure the theoretical framework which could serve as a basis for assessing all future applications of Aligned Structuring models. Second, while many scholars have chosen to examine Aligned Structuring models by analogy or comparison to governance challenges encountered by publicly traded firms[59] or nonprofit enterprises,[60] this paper emphasizes OpenAI and Anthropic's nature as technological startups. Analyzing their novel business structures with the specialized toolset developed by startup law scholars over the last decades is crucial for understanding the arrays of incentives that operate on AI startups, their shareholders, and their stakeholders accurately.

This paper proceeds in the following order. Part I provides a concise review of the basic concepts of "AI safety" and "AI ethics" and of OpenAI and Anthropic's business structures. Part II highlights how these structures deviate from standard startup structuring practices and how nontrivial these deviations are. Part III discusses separately each of the three novelties in OpenAI and Anthropic's business structures, systematically exploring their capabilities, weak spots, and costs. Part IV weighs the benefits and costs of Aligned Structuring models based on the discussion in the preceding parts. A brief conclusion follows.

I. ALIGNED STRUCTURING BASICS

A. Approaches to AI Risk

1. AI Safety

Within the broad spectrum of potential risks associated with the development and use of AI technology, the term "*AI Safety*" (or sometimes, "*AI Alignment*")[61] is typically associated with the discussion of future risks

---

58. *See, e.g.*, Andhov, *supra* note 56; Pedersen, *supra* note 57; Trautman & Foster, II, *supra* note 57.

59. *See, e.g.*, Gatti, *supra* note 56; Tallarita, *supra* note 19.

60. *See, e.g.*, Cathy Hwang & Dorothy S. Lund, *Purpose and Nonprofit Enterprise* (Eur. Corp. Gov. Inst., Working Paper No. 819, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5061739 [https://perma.cc/57J4-H25F]; Pedersen, *supra* note 57.

61. The term "Alignment" may be used interchangeably with the term "AI safety," see Sam Bowman, *AI Safety and Neighboring Communities: A Quick-Start Guide, as of Summer 2022*, AI ALIGNMENT F. (Sept. 1, 2022), https://www.alignmentforum.org/posts/EFpQcBmfm2bFfM4zM/

of catastrophic magnitude.[62] A key concern for AI safety advocates is the premature arrival of super-intelligent machines with *Artificial General Intelligence* (AGI)—that is, the ability to outperform humans across a wide variety of tasks.[63] Unless proper controls are established in their design – a process described by safety researchers as "*alignment"*[64]—AI or AGI-powered machines might defy their humble human operators' control and seriously undermine human interests and values—perhaps even humanity's mere existence.[65] Such scenarios are referred to in AI safety jargon as "*existential risks*" (or "*x-risks*").[66]

AI safety theorists describe how machines' counter-human behavior might result from three core problems. First, humans might fail to define the machine's mission properly.[67] Second, super-intelligent machines might seek power and control on their own initiative to fulfill their missions more efficiently (an *instrumental convergence* problem).[68] Lastly, machines do not naturally share or appreciate the interests and values that humans typically care about and, therefore, unless ordered otherwise, may disregard them if standing in the way of fulfilling their mission effectively (an *orthogonality*

---

ai-safety-and-neighboring-communities-a-quick-start-guide-as [https://perma.cc/G6E6-9H6E], or to describe the subfield of AI safety focusing on how dangerous AI behavior could be effectively prevented. Shazeda Ahmed et al., *Building the Epistemic Community of AI Safety*, SSRN (Dec. 1, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4641526 [https://perma.cc/4BRE-W6P9] (manuscript at 6); Aaron J. Snoswell, *What Is 'AI Alignment'? Silicon Valley's Favourite Way to Think About AI Safety Misses the Real Issues*, CONVERSATION (July 11, 2023), https://theconversation.com/what-is-ai-alignment-silicon-valleys-favourite-way-to-think-about-ai-safety-misses-the-real-issues-209330 [https://perma.cc/MLG5-HZPY] (noting that "many in Silicon Valley view safety through the speculative lens of "AI alignment").

62. Ahmed et al., *supra* note 61, at 4; Bowman, *supra* note 61 (defining AI safety as "[t]he research project of ensuring that future AI progress doesn't yield civilization-endingly catastrophic results"). *But see infra* note 71 for information about the gradual expansion of the term.

63. For information about the definition of AGI, see Bowen Xu, What Is Meant by AGI? On the Definition of Artificial General Intelligence, §§ 3.4–4 (Apr. 16, 2024) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2404.10731 [https://perma.cc/W7XN-4J6P].

64. *See supra* note 61.

65. *See generally* Yonathan A. Arbel et al., *Systemic Regulation of Artificial Intelligence*, 56 ARIZ. ST. L.J. 545 (2024); *see also* Bowman, *supra* note 61 ("Once you have a system with superhuman reasoning and planning abilities, it's easy to make it dangerous by accident . . . If such a system were single-mindedly pursuing a dangerous goal, we probably wouldn't be able to stop it.").

66. Ahmed et al., *supra* note 61, at 3–4.

67. Arbel et al., *supra* note 65, at 26–28.

68. *Id.* at 28–29; Nick Bostrom, *The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents*, 22 MIND & MACHS. 71, 76–83 (2012).

problem).[69] While AI safety risks might seem to be taken straight from science fiction, many AI researchers believe they are likely to pose a serious threat to humankind within a few years to a few decades.[70]

2. AI Ethics

Although the distinction is neither absolute nor universally agreed upon,[71] the term "AI *ethics*" (or sometimes "*Responsible AI*") is often used to describe the AI risk discourse that focuses on tangible risks that AI technologies may pose at present or in the near future.[72] AI ethics advocates raise concerns about various societal and moral issues stemming from training and using AI models, from AI-generated hate speech and propaganda on social media to AI-related job displacement to the environmental impact of the supercomputers running complex AI models.[73] The ethics discourse, however, is often described as being focused on two core subjects: *AI bias* and *AI privacy*.[74] AI bias concerns relate to the observation that AI models

---

69. Arbel et al., *supra* note 65, at 26–28; Bostrom, *supra* note 68, at 71–76.

70. Arbel et al., *supra* note 65, at 31; Bowman, *supra* note 61 ("Consensus within the field is that these risks could become concrete within ~4–25 years, and have a >10% chance of being leading to a global catastrophe (i.e., extinction or something comparably bad."); Tallarita, *supra* note 19 ("The consensus among AI researchers is that human-level AI is imminent, and the alignment problem is real.").

71. Reed Albergotti, *The Risks of Expanding the Definition of 'AI Safety*,*'* SEMAFOR (Mar. 8, 2024), https://www.semafor.com/article/03/08/2024/the-risks-of-expanding-the-definition-of-ai-safety [https://perma.cc/3XU8-PLX2]; Bowman, *supra* note 61 (commenting that "lots of 'AI Safety' projects/funders have little to do" with core safety issues such as AI alignment).

72. Bowman, *supra* note 61 (defining AI ethics as "the research and political project of minimizing the harms of current and near-future AI/ML technology and of ensuring that any benefits from such technology are shared broadly"); Kelsey Piper, *There Are Two Factions Working to Prevent AI Dangers. Here's Why They're Deeply Divided*, VOX (Aug. 10, 2022), https://www.vox.com/future-perfect/2022/8/10/23298108/ai-dangers-ethics-alignment-present-future-risk [https://perma.cc/5ZRG-P5Y3]; Henrik Skaug Sætra & John Danaher, *Resolving the Battle of Short- vs. Long-term AI Risks*, AI & ETHICS (Sept. 4, 2023), https://link.springer.com/article/10.1007/s43681-023-00336-y [https://perma.cc/6SZ3-93X9], at § 4.

73. *See generally* Chris Daily, *Define Ethics: An Introduction to AI and Its Implications*, MEDIUM (Oct. 23, 2023), https://medium.com/the-ai-educator/define-ethics-an-introduction-to-ai-and-its-implications-a19c2096bcda; Bruce Schneier & Nathan Sanders, *The A.I. Wars Have Three Factions, and They All Crave Power*, N.Y. TIMES (Sept. 28, 2023), https://nyti.ms/4azCAZw (under "The Reformers").

74. *See, e.g.*, Kif Leswing, *Parrots, Paper Clips and Safety vs. Ethics: Why the Artificial Intelligence Debate Sounds Like a Foreign Language*, CNBC (May 30, 2023), https://cnb.cx/3QCkhMi [https://perma.cc/7SAZ-VYGW]; Piper, *supra* note 72; Aryan Saxena et al., *The Ethics of AI: Addressing Bias and Privacy Concerns*, MEDIUM (June 21, 2023), https://medium.com/@IEEE_Computer_Society_VIT/the-ethics-of-ai-addressing-bias-and-privacy-concerns-573df8ca1490.

might inherit societal prejudices reflected in their training data—notably, the underrepresentation of disadvantaged populations.[75] AI privacy concerns, on the other hand, emphasize the unprecedented data collection and analysis capabilities of advanced AI and their implications on data subjects' rights. AI privacy research attempts to develop data anonymization techniques, establish data governance frameworks, and advocate for stricter regulations on data collection and usage practices.[76]

B. OpenAI and Anthropic's Aligned Structuring Models

AI safety and ethics concerns may be addressed in various ways. One strategy would be to rely on direct regulatory oversight—establishing and enforcing binding rules that dictate permissible and impermissible activities for AI developers.[77] An alternative path would be to focus on governance mechanisms entrusted to AI developers and their internal rulemaking and oversight. These typically include setting internal risk management policies, ethical guidelines, codes of conduct, training programs, or safety committees focused on AI risks within organizations engaging in AI development.[78]

As this paper illustrates, Aligned Structuring represents a different strategy for addressing AI risks through self-governance. Instead of using direct means that are closer to the startup's day-to-day operations—such as setting ethical guidelines, safety committees, or compliance systems—Aligned Structuring seeks to foster safe and ethical conduct more indirectly, through reshaping the foundations of the startup's business structure: its

---

75. See generally Emilio Ferrara, *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies*, 6 SCI. 3 (2024), for a discussion of the root causes of AI bias and ways to address them; Mi Zhou et al., Bias in Generative AI (Mar. 5, 2024) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2403.02726 [https://perma.cc/9TFQ-L72Y] (finding, for example, that AI image generators Midjourney, Stable Diffusion, and DALL-E 2 exhibit biases against women and African Americans).

76. *See generally* Hao-Ping (Hank) Lee et al., Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks (Feb. 10, 2024) (unpublished manuscript) (on file with arXiv), https://arxiv.org/abs/2310.07879 [https://perma.cc/8T8M-D8KN] (analyzing a dataset of documented AI privacy incidents and offering a classification of AI privacy risks).

77. A notable example of such efforts is the newly adopted EU AI Act. *See generally* Martin Kretschmer et al., The Risks of Risk-Based AI Regulation: Taking Liability Seriously § 2 (Nov. 28, 2023) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2311.14684 [https://perma.cc/8BDG-9X6G]; Jonas Tallberg et al., *AI Regulation in the European Union: Examining Non-State Actors' Preferences*, 26 BUS. & POL. 218, 218–19 (2024).

78. See, for instance, the internal AI governance standards adopted by Microsoft (MICROSOFT RESPONSIBLE AI STANDARD, V2 – GENERAL REQUIREMENTS (2022)) and Google (GOOGLE, AI PRINCIPLES PROGRESS UPDATE (2023)).

choice of entity and its fundamental allocation of control and economic rights between owners, managers, and stakeholders.

OpenAI and Anthropic, the two startups that have famously applied Aligned Structuring models so far, have more than that in common. Notably, they are both developing their own *foundation models*—that is, massive, general-purpose AI models which may serve as a platform for developing specialized AI applications tailored to specific industries or tasks.[79] These startups have been known to require enormous investments of financial capital and computing power, much more than is traditionally associated with Silicon Valley startups.[80] This fact, as will be further discussed below, has dramatic implications for such startups' relationships with their stakeholders and their potential to achieve their intended purpose of aligning startup stakeholders' incentives with those of society as a whole.

This section will review the unique features of OpenAI and Anthropic's Aligned Structuring models. It will first provide an outline of these structures and proceed to analyze how they differ from startup structuring conventions. The following sections will dig deeper into the reasoning behind these changes and how they are claimed to serve their designated purpose.

1. OpenAI's Capped Profit Structure

OpenAI—best known as the maker of the large language model series GPT[81]—was founded in 2015 as a nonprofit research lab on a mission to "advance digital intelligence in the way that is most likely to benefit humanity as a whole."[82] Its stellar list of entrepreneurs included, among others, Elon Musk of Tesla and SpaceX; Sam Altman, formerly President of the well-known startup accelerator Y Combinator; Greg Brockman, former CTO of payment processing unicorn Stripe; and renowned ex-Google computer scientist Ilya Sutskever.[83] Among its early supporters, collectively committing $1 billion in donations, were tech giants Amazon Web Services and Infosys and Silicon Valley icons Reid Hoffman, Jessica Livingston, and Peter Thiel.[84]

---

79. Matthew R. Gaske, *Regulation Priorities for Artificial Intelligence Foundation Models*, 26 VAND. J. ENT. & TECH. L. 1, 5–6 (2023).

80. Kak et al., *supra* note 16; Riekeles & von Thun, *supra* note 16.

81. Yutong Levine, *A Timeline of OpenAI's Technology, Funding, and History*, MEDIUM (Nov. 20, 2023), https://medium.com/@DiscoverLevine/a-timeline-of-openais-technology-funding-and-history-c91cbc071a85.

82. *Introducing OpenAI*, *supra* note 13.

83. *Id.*

84. *Id.*

In March 2019, OpenAI announced on its website a dramatic change in its business model—a transition from a pure nonprofit into a bespoke "*capped profit*" structure.[85] Assuring its followers that its mission remained "to ensure that artificial general intelligence (AGI) benefits all of humanity, primarily by attempting to build safe AGI and share the benefits with the world," OpenAI acknowledged that to stay on the wheel, it must grow faster than initially anticipated.[86] This, reportedly, required "invest[ing] billions of dollars in upcoming years into large-scale cloud compute, attracting and retaining talented people, and building AI supercomputers."[87] The new structure, marketed as "a hybrid between a for-profit and nonprofit," was designed to strike "the right balance," allowing OpenAI to raise capital like a for-profit startup while staying true to its nonprofit mission.[88] Shortly after that, OpenAI entered into a strategic partnership with Microsoft, which involved licensing OpenAI's technology to Microsoft, coupled with a series of multibillion-dollar investment commitments by the tech titan over the following years.[89]

As of August 2024, OpenAI's structure was described on OpenAI's website as follows.[90] An LLC (OpenAI Global) was formed and used to raise capital from investors and hire employees who work on for-profit projects.[91] While a for-profit entity, OpenAI Global was fully controlled—indirectly, and through intermediate entities—by the original nonprofit parent (OpenAI, Inc.)[92] and was "legally bound to pursue the Nonprofit's mission."[93] Effectively, OpenAI Global was controlled by the nonprofit parent's board of directors.[94]

OpenAI has taken several measures to protect the nonprofit board's independence from investor pressure and profit-seeking incentives. The

---

85. *Capped Profit Announcement*, *supra* note 24.

86. *Id.*

87. *Id.*

88. *Id.*

89. OpenAI noted that "AGI is explicitly carved out of all commercial and IP licensing agreements" with Microsoft. *Capped Profit Structure Primer*, *supra* note 13. As of 2024, Microsoft has reportedly committed to invest over $13 billion in OpenAI. Andrew Ross Sorkin et al., *The F.T.C. Takes on A.I. Deals*, N.Y. TIMES (Jan. 26, 2024), https://nyti.ms/3Kc00tb.

90. *Capped Profit Structure Primer*, *supra* note 13. The capped profit structure seems to have gone through some changes between being first introduced and finally replaced. *See* Musk v. Altman, 2024 WL 899024, §§ 70–76 (Cal. Super.); Pedersen, *supra* note 57, at § 4.

91. *Id.* The nonprofit parent held OpenAI Global's equity indirectly, through a holding company. *See infra* note 99.

92. The nonprofit's control over the for-profit was indirect and facilitated through OpenAI GP LLC, a "manager entity" controlled by the for-profit. *See id.*

93. *Id.*

94. *See* Pedersen, *supra* note 57, n.3.

board maintained a majority of so-called "independent directors" who hold no OpenAI equity.[95] All nonprofit parent's directors, independent or not, were required to "perform their fiduciary duties in furtherance of its mission—safe AGI that is broadly beneficial."[96] The nonprofit's board was given, at least at some point, the exclusive right to appoint and remove fellow directors and determine the board size.[97] Microsoft was a non-voting board observer.[98]

The allocation of OpenAI Global's economic rights—which likely included the right to participate in future distributions of profits and liquidation proceeds—was the heart of the capped-profit structure. OpenAI Global's economic rights were shared among its investors, employees, and nonprofit parent.[99] However, investors and employees may have only participated in distributions up to a certain cap. Any "residual value created above and beyond the cap" was designated to flow back to the nonprofit parent "for the benefit of humanity."[100] OpenAI disclosed that returns for its first round of investors were capped at 100x (up to 100 times their original investment).[101] Later investors were rumored in the media to have been subject to lower caps.[102]

---

95. *Capped Profit Structure Primer*, *supra* note 13.

96. *Id.*

97. Provisions in this respect were reportedly included in OpenAI's 2016 bylaws, but it is unclear whether and to what extent they remained in force throughout OpenAI's capped-profit stage. *See* Paresh Dave, *How OpenAI's Bizarre Structure Gave 4 People the Power to Fire Sam Altman*, WIRED (Nov. 19, 2023), https://www.wired.com/story/openai-bizarre-structure-4-people-the-power-to-fire-sam-altman/; Pedersen, *supra* note 57, § 4.4. After its recent restructuring, *see* discussion *infra* notes 110–11 and accompanying text, the nonprofit board appoints all members of the for-profit entity, *Our Structure*, OPENAI https://openai.com/our-structure (last visited Feb. 17, 2026).

98. *Capped Profit Structure Primer*, *supra* note 13. Microsoft reportedly later relinquished its seat as an observer. *See* Mauro Orru & Christian Moess Laursen, *Microsoft Quits OpenAI's Board Amid Antitrust Scrutiny*, WALL ST. J. (July 10, 2024), https://on.wsj.com/3ywg9Yw. For information about board observers, see generally Nizan Geslevich Packin & Anat Alon-Beck, *Board Observers*, SSRN (July 25, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4745278 [https://perma.cc/VPL9-PHQ2].

99. Microsoft's economic rights seem to have been held directly, while the nonprofit parent, employees, and non-Microsoft investors' economic rights were held indirectly through a holding company. *See Capped Profit Structure Primer*, *supra* note 13.

100. *See Capped Profit Structure Primer*, *supra* note 13.

101. *Capped Profit Announcement*, *supra* note 24.

102. Tima Bansal, *Does OpenAI's Non-Profit Ownership Structure Actually Matter?*, FORBES (Oct. 13, 2023), https://www.forbes.com/sites/timabansal/2023/10/13/does-openais-non-profit-ownership-structure-actually-matter/?sh=7871dbe97d18 ("Under the new structure, investors' profits would be capped at 100 times their investment. . . . Later investors received much smaller profit multipliers. For example, Microsoft's $1-billion investment in 2019 was rumored to be capped at 20 times.").

In late 2024, OpenAI proposed a major restructuring as part of a multibillion-dollar fundraising round that would have converted OpenAI Global into a PBC and replaced the nonprofit parent's complete control over OpenAI Global with a "significant interest" in "the form of shares in the PBC."[103] Another purported aspect of the new structure was rolling back the capping of at least some equity holders' profits.[104] The plan, which was reported to have been driven by investor expectations,[105] sparked fierce backlash from AI researchers, policy leaders, and former insiders, prompting some of them to publish an open letter to the Attorneys General of Delaware and California, titled "Not for Private Gain," warning that the restructuring would eliminate the nonprofit guardrails distinguishing OpenAI from conventional tech firms.[106] After weeks of public criticism and internal debate, and with a legal action by Elon Musk to block the move hovering above,[107] OpenAI revised the proposal in May 2025.[108] The new proposal maintained the PBC transition but preserved some extent of nonprofit control over it, while, reportedly, still removing the profit caps off equity holders' stakes.[109]

In October 2025, OpenAI announced that its restructuring was complete, with the nonprofit entity (now "OpenAI Foundation") maintaining control over the for-profit entity (now organized as a PBC and named "OpenAI Group").[110] The announcement also mentioned that following the

---

103. *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/ [https://perma.cc/BLR7-SN5T].

104. *Not for Private Gain First Letter*, *supra* note 1; *see also* Aditya Soni et al., *OpenAI Outlines New For-profit Structure in Bid to Stay Ahead in Costly AI Race*, REUTERS (Dec. 27, 2024), https://www.reuters.com/technology/artificial-intelligence/openai-lays-out-plan-shift-new-for-profit-structure-2024-12-27/ ("[OpenAI's] latest $6.6 billion funding round . . . was contingent on whether the ChatGPT-maker could upend its corporate structure and remove a profit cap for investors within two years.").

105. *Not for Private Gain First Letter*, *supra* note 1 ("OpenAI recently announced that it intends to restructure again, purportedly to satisfy investor demands to simplify the capital structure.").

106. *Id.*

107. Musk v. Altman, 4:24-cv-04722-YGR (N.D. Cal. Nov. 29, 2024).

108. *Evolving OpenAI's Structure*, OpenAI (May 5, 2025), https://openai.com/index/evolving-our-structure/ [https://perma.cc/Z8EB-ZGZF].

109. *Id.*; *see also* Page Hedley et al., *May 12 Letter*, NOT FOR PRIVATE GAIN (May 12, 2025), https://notforprivategain.org/follow-up [https://perma.cc/2SRS-TCHV] [hereinafter *Not for Private Gain Second Letter*]

110. Bret Taylor, *Built to Benefit Everyone*, OPENAI (Oct. 28, 2025), https://openai.com/index/built-to-benefit-everyone [https://perma.cc/LP86-9KZH]; *Our Structure*, *supra* note 97. For discussion of the extent of control the nonprofit entity has in fact over the for-profit entity under the new structure, see Page Hedley et al., *Nov 3 Update*, NOT FOR PRIVATE GAIN (Nov. 3,

restructuring, "all equity holders in OpenAI Group now own the same type of traditional stock that participates proportionally and grows in value with OpenAI Group's success," which could be read as suggesting a rollback or changes in the nature of the profit caps.[111] The Attorneys General of Delaware and California have indicated they will not object the restructuring on the terms presented to them.[112]

2. Anthropic's Long-Term Benefit Trust

Anthropic—best known as the creator of the large language model series Claude[113]—was founded in 2021 by a group of former OpenAI employees led by Dario Amodei.[114] Their departure was attributed in media reports to mission drift concerns relating to OpenAI's balancing of safety and profitability.[115]

Anthropic's early financing rounds were remarkable in their sheer size, proximity, and being led by known advocates of x-risk awareness rather than

---

2025), https://notforprivategain.org/november-update [https://perma.cc/9RWV-BXND] [hereinafter *Not for Private Gain Third Letter*].

111. *Our Structure*, *supra* note 97; *see also* Celia Ford, *What You Need to Know About the OpenAI Restructure*, TRANSFORMER (Oct. 29, 2025), https://www.transformernews.ai/p/what-you-need-to-know-about-the-openai-restructure-sam-altman-pbc-foundation [https://perma.cc/DBP9-TXHW] ("now, the profit caps are gone"). *But see Not For Private Gain Third Letter*, *supra* note 110 ("OpenAI . . . has not disclosed the details of the current profit caps").

112. *AG Jennings Completes Review of OpenAI Recapitalization*, DELAWARE NEWS (Oct. 28, 2025), https://news.delaware.gov/2025/10/28/ag-jennings-completes-review-of-openai-recapitalization [https://perma.cc/5GS3-P962]; *Attorney General Bonta Issues Statement on OpenAI's Recapitalization Plan*, ROB BONTA ATTORNEY GENERAL (Oct. 28, 2025), https://oag.ca.gov/news/press-releases/attorney-general-bonta-issues-statement-openai's-recapitalization-plan [https://perma.cc/B7F2-F3ML].

113. Kyle Wiggers, *Anthropic Claims Its New AI Chatbot Models Beat OpenAI's GPT-4*, TECHCRUNCH (Mar. 4, 2024), https://tcrn.ch/44Yklfo [https://perma.cc/E2LH-Y5E7].

114. Richard Waters & Miles Kruppa, *Rebel AI Group Raises Record Cash After Machine Learning Schism*, FIN. TIMES (May 28, 2021), https://www.ft.com/content/8de92f3a-228e-4bb8-961f-96f2dce70ebb.

115. *Id.*; *see also* Dan Holden, *Anthropic's Sibling Founders on Leaving OpenAI to Start a $15B Startup*, OBSERVER (May 10, 2024), https://observer.com/2024/05/anthropics-sibling-founders-on-leaving-openai-to-start-a-15b-startup/ [https://perma.cc/F6FK-K9GX] (quoting Daniela Amodei: "We left OpenAI because of concerns around the direction. . . . [W]e wanted to be sure the tools were being used reliably and responsibly . . . We want to be the most responsible A.I. we can, always asking the question, 'What could go wrong here?'"); Lakshmi Varanasi, *Dario Amodei Says Anthropic Struggles to Balance 'Incredible Commercial Pressure' with Its 'Safety Stuff'*, BUS. INSIDER (Feb. 16, 2026), https://www.businessinsider.com/dario-amodei-anthropic-profit-pressure-versus-safety-mission-2026-2 [https://perma.cc/PH9P-AQSF].

traditional early-stage startup investors.[116] Less than three years from its formation and with a headcount of roughly 40 people, Anthropic had already raised a $124 million Series A led by Skype co-founder Jaan Tallinn[117] and a jaw-dropping $580 million Series B led by FTX's Sam Bankman-Fried.[118]

In mid-2023, Anthropic announced the closing of its $450 million Series C—led by the VC fund Spark Capital—with participation from Google, Salesforce, Zoom, and others.[119] A few months later, Anthropic revealed that the Series C involved adopting a new business structure, which was marketed as a means to help "align" the firm's governance with its "mission of developing and maintaining advanced AI for the long-term benefit of humanity."[120]

The new structure was described by Anthropic and some of its outside architects[121] to operate as follows. The structure's base layer relies on Anthropic's classification as a Delaware PBC.[122] As discussed below,[123] unlike the directors of standard corporations, PBC directors must balance stockholders' financial interests, broader stakeholders' interests, and the public benefit identified in their charter when exercising their judgment.[124] The specific public benefit that Anthropic's directors must consider, as

---

116. Erin Griffith & Cade Metz, *Inside the Funding Frenzy at Anthropic*, *One of A.I.'s Hottest Start-Ups*, N.Y. TIMES (Feb. 20, 2024), https://nyti.ms/3QTOd6v.

117. *Anthropic Raises $124 Million to Build More Reliable, General AI Systems*, ANTHROPIC (May 28, 2021), https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems [https://perma.cc/7W7Q-LTAS]. According to Carta data, the median Series A deal size in the second quarter of 2021 was $11 million. Kevin Dowd & Peter Walker, *State of Private Markets: Q4 and 2022 in Review*, CARTA (Feb. 10, 2023), https://carta.com/data/state-of-private-markets-q4-2022/ [https://perma.cc/X7Y3-4H6M] [hereinafter *State of Private Markets 2022*].

118. *Anthropic Raises Series B to Build Steerable, Interpretable, Robust AI Systems*, ANTHROPIC (Apr. 29, 2022), https://www.anthropic.com/news/anthropic-raises-series-b-to-build-safe-reliable-ai [https://perma.cc/R5CY-AXGL]. According to Carta data, the median Series B deal size in the second quarter of 2022 was $23 million. Dowd & Walker, *supra* note 117.

119. *Anthropic Raises $450 Million in Series C Funding to Scale Reliable AI Products*, ANTHROPIC (May 23, 2023), https://www.anthropic.com/news/anthropic-series-c [https://perma.cc/ZT3W-JVK5]. According to Carta data, the median Series C deal size in the second quarter of 2023 was $25 million. Kevin Dowd, *Startup Round Sizes Got Bigger at Most Stages in Q1*, CARTA (Apr. 17, 2024), https://carta.com/uk/en/data/round-sizes-q1-2024/ [https://perma.cc/V26X-3F36].

120. *LTBT Primer*, *supra* note 6.

121. Anthropic acknowledged on its announcement of the LTBT structure that the LTBT "was designed and 'red teamed' with immeasurable assistance by John Morley of Yale Law School, David Berger, Amy Simmerman, and other lawyers from Wilson Sonsini, and by Noah Feldman and Seth Berman from Harvard Law School and Ethical Compass Advisors." *Id.*

122. *Id.*

123. *See infra* notes 295–96 and accompanying text.

124. DEL. CODE ANN. tit. 8, § 365(a) (West 2020).

identified in its charter, is to "responsibly develop and maintain advanced AI for the long-term benefit of humanity."[125]

The newly added second layer was based on the allocation of powerful control rights in Anthropic to the "Long-Term Benefit Trust" (LTBT), a Delaware purpose trust created outside Anthropic.[126] The LTBT holds a special class of Anthropic's stock that comes with a preferential set of control rights.[127] These include, primarily, the right to elect a gradually increasing number of Anthropic's board members—and eventually, the board majority.[128] The LTBT was also granted a set of "protective provisions" over "certain actions that could significantly alter the corporation or its business."[129]

Unlike classic common law trusts, which are created for the benefit of one or more persons, purpose trusts like the LTBT may be settled—where allowed by statute—with a mandate to promote a designated purpose or cause.[130] The LTBT's particular purpose was designed to mirror Anthropic's abovementioned public purpose.[131] The LTBT's trustees—like Anthropic's directors—are legally obligated to exercise their control and judgment while balancing between Anthropic shareholders' pecuniary interests, the interests of the broader circle of stakeholders materially affected by Anthropic's conduct, and Anthropic's commitment to responsible AI development.[132]

The LTBT comprises five trustees possessing "backgrounds and expertise in AI safety, national security, public policy, and social enterprise."[133] To protect the LTBT's independent and unbiased judgment, all trustees are reportedly "financially disinterested."[134] Furthermore, LTBT trustees serve one-year terms and are elected by the trustees then-serving—although Anthropic itself chose the initial trustees.[135] However, the LTBT's independence is somewhat constrained by a so-called "failsafe" mechanism, allowing "sufficiently large supermajorities" of Anthropic's shareholders to make changes to the trust and its powers without the trustees' consent.[136]

---

125. *LTBT Primer*, *supra* note 6.

126. *Id.*

127. *Id.* The Class T Common Stock reportedly does, however, represent some "very small" financial claim—perhaps their par value. Morley et al., *supra* note 6.

128. *LTBT Primer*, *supra* note 6.

129. *Id.*

130. Morley et al., *supra* note 6.

131. *Id.*

132. *Id.*

133. *LTBT Primer*, *supra* note 6.

134. *Id.*

135. *Id.*

136. *Id.*

II. ALIGNED STRUCTURING V. THE STARTUP STRUCTURING PLAYBOOK

OpenAI's capped profit structure and Anthropic's LTBT represent significant departures from conventional startup structuring. This section will provide an overview of these deviations, and the following sections will dig deeper into their apparent motivations.

A. Control Rights

Startups are typically governed by their entrepreneurs and investors, with their governance structures following a predictable pattern in which control gradually shifts from entrepreneurs to investors over time.[137] Each VC leading a major investment round in a startup would usually secure a board seat by increasing the board size or having an entrepreneur or a previous investor forgo their seat.[138] Smaller investors may secure a board observer or information rights, allowing them some visibility into the board's actions.[139] And so, as more investment rounds are raised, the startup's board would usually shift from undisputed entrepreneur control at inception to a shared control structure—having an equal number of directors appointed by entrepreneurs and investors.[140] In such cases, tie-breaking directors are often appointed, typically at the entrepreneurs' and investors' mutual consent or by the vote of the shareholders as a whole.[141] At the startup's later stages, investors may end up holding a majority of the startup's board seats.[142] At the same time, as more equity is issued to investors with each investment round, investors' shareholder-level voting power increases, and entrepreneurs' voting power is gradually diluted.[143] Startup investors would often also hold a set of veto rights over strategic corporate decisions, allowing them a significant extent of control even before having a board or shareholder-level majority.[144]

---

137. *Pollman*, *supra* note 11, at 181.

138. FELD & MENDELSON, *supra* note 11, at 77–81; Pollman, *supra* note 11, at 181; Smith, *supra* note 11, at 326–27.

139. Broughman & Wansley, *supra* note 42, at 1312; FELD & MENDELSON, *supra* note 11, at 79.

140. FELD & MENDELSON, *supra* note 11, at 77–81; Pollman, *supra* note 11, at 181; Smith *supra* note 11, at 326–27.

141. FELD & MENDELSON, *supra* note 11, at 79–80; Pollman, *supra* note 11, at 181; Smith, *supra* note 11, at 326–27.

142. Pollman, *supra* note 11, at 181; Smith, *supra* note 11, at 326–27.

143. Broughman & Wansley, *supra* note 42, at 1312; Pollman, *supra* note 11, at 181.

144. Broughman & Wansley, *supra* note 42, at 1312; Pollman, *supra* note 11, at 181 n.144; Smith, *supra* note 11, at 346–47.

Startup scholarship of the last three decades explored how allocating control rights between entrepreneurs and investors is critical in establishing efficient governance structures. Notably, startup control rights bridge *information asymmetries* between entrepreneurs and investors. Due to the novelty and complexity of startups' technologies and the entrepreneurs' lack of track record, entrepreneurs typically possess much more information than investors about the technology's true potential and their capabilities as managers.[145] As a result, it is difficult for startup investors to distinguish between promising and less promising investment opportunities, hindering their ability to adjust their term sheets accordingly.[146] Investor control has often been described as a means to overcome these information asymmetries. For instance, active involvement in the startup's operations allows investors to collect valuable information on the product and entrepreneurs and use it to make informed decisions when making offers for follow-up investments.[147]

The allocation of control is also used to minimize *agency costs* within startups' governance structures. Imagine, for example, an early-stage startup where the entrepreneurs hold both board control and majority ownership. Here, the entrepreneurs primarily act as agents, while investors are the principals.[148] As mentioned, non-controlling startup investors often hold veto power over significant business decisions.[149] These rights reduce agency costs by protecting investors from managerial decisions resulting from self-dealing[150] or honest mistakes.[151] To illustrate, investors might veto a proposed

---

145. Broughman & Wansley, *supra* note 42, at 1307–08; Gilson, *supra* note 15, at 1076–78, 1080; GOMPERS & LERNER, *supra* note 15, at 3.

146. Broughman & Wansley, *supra* note 42, at 1307–08; Gilson, *supra* note 15, at 1076–78, 1080; GOMPERS & LERNER, *supra* note 15, at 3.

147. Broughman & Wansley, *supra* note 42, at 1312–13; Gilson, *supra* note 15, at 1076–78, at 1082–83; GOMPERS & LERNER, *supra* note 15, at 160; Sahlman, *supra* note 15, at 508–09.

148. Pollman, *supra* note 11, at 160; *see also* D. Gordon Smith, *Venture Capital Contracting in the Information Age*, 2 J. SMALL & EMERGING BUS. L. 133, 138–39 (1998) (discussing that startup entrepreneurs and investors typically both act in some principal and agent capacity, and therefore, startup governance is sometimes described as a nexus of agency relationships).

149. *See* Smith, *supra* note 11.

150. The costs of agent self-serving are referred to as "agency costs" in the traditional model developed by Jensen and Meckling. Michael C. Jensen & William H. Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, 3 J. FIN. ECON. 305, 308–10 (1976). Also, "agent conflict costs" under the "principal costs" model were developed by Zohar Goshen and Richard Squire. Goshen & Squire, *supra* note 20, at 793–94. For information about the use of investor control and monitoring to minimize entrepreneur-side conflict costs see Broughman & Wansley, *supra* note 42, at 1313; and Gilson, *supra* note 15, at 1082; GOMPERS & LERNER, *supra* note 15, at 160–61.

151. The costs of agents' honest mistakes are referred to as "agent competence costs" in Goshen and Squire's model. Goshen & Squire, *supra* note 20, at 788. For a discussion about the

financing round if they suspect that the entrepreneurs are unreasonably biased towards the VC that funded their previous startup or that the inexperienced entrepreneurs are too anxious to close and overlook better offers.

Entrepreneurs—considering the extent of veto rights they should allow their investors to hold—also consider the potential inefficiencies and costs resulting from investor opportunism or insufficient skill. For instance, self-serving investors might veto a new financing round with favorable terms because they want to lead it themselves.[152] Investors with an inadequate understanding of the technology and its target market might do so because they fail to appreciate the strategic value of the new investor and its potential contribution to developing or commercializing the startup's product.[153]

Therefore, the extent and nature of control rights offered to investors are expected to factor in the level of trust between the two parties and their mutual attempts to reduce agency costs.[154] Investors who believe in the entrepreneurs' good intentions and sound judgment will likely require fewer controls. Conversely, a lack of trust on the part of the entrepreneurs may lead them to resist granting extensive control rights to their investors.

Both OpenAI and Anthropic have broken the mold by creating startup governance structures where the allocation of control is no longer a zero-sum game between entrepreneurs and investors. Instead, both startups have chosen to allocate a significant extent of control to their respective Alignment Stewards, who are presumed to be isolated from financial motivations and represent society's interest in ensuring AI technology is safe and ethical. In Anthropic, this position is filled by the LTBT—an independent third party, presumably having no ties to Anthropic's entrepreneurs or investors, that holds protective provisions and will eventually hold the power to elect the

---

use of investor monitoring to mitigate the costs of entrepreneurs' honest business mistakes see, for example, Fried & Ganor, *supra* note 42, at 989–90.

152. The costs of principal opportunism are referred to as "principal conflict costs" in Goshen and Squire's model. Goshen & Squire, *supra* note 20, at 791–93. For a discussion about the tradeoff between investor control and principal conflict costs in startups see, for example, Gilson, *supra* note 15, at 1086–87; Robert P. Bartlett, III, *Venture Capital, Agency Costs, and the False Dichotomy of the Corporation*, 54 UCLA L. REV. 37, 56 n.78 (2006).

153. The costs of principal incompetence are referred to as "principal competence costs" in Goshen and Squire's model. Goshen & Squire, *supra* note 20, at 786–88. For a discussion about investors' information disadvantage on product-related matters, see *supra* note 145 and accompanying text; and *infra* notes 234–48 and accompanying text.

154. That is not to say that agency costs are the *only* factor determining the allocation of startups' control rights; bargaining power differences between entrepreneurs and investors may also have an effect. *See generally* Spencer Williams, *Venture Capital Contract Design: An Empirical Analysis of the Connection Between Bargaining Power and Venture Financing Contract Terms*, 23 FORDHAM J. CORP. & FIN. L. 105 (2017) (finding that the market availability of VC funding significantly impacts both the price and non-price terms of VC deals).

majority of Anthropic's board.[155] In OpenAI, the nonprofit entity, the Alignment Steward, exercises complete control over the for-profit entity (previously OpenAI Global, and currently OpenAI Group PBC) —while being governed by a board with a majority of independent directors, not by OpenAI's entrepreneurs or investors.[156]

To be sure, Alignment Stewards are fundamentally different from tie-breaking directors—the more common type of non-entrepreneur, non-investor members who may populate startups' boards. While Alignment Stewards are presumed independent of the startup and its owners,[157] tie-breaking directors unquestionably *depend* on the startup's entrepreneurs' and investors' approval and only hold office at their pleasure.[158] They are appointed to help resolve deadlocks and disagreements among profit-seeking shareholders, not to represent broader stakeholders' interests.[159] If they enjoy any independence at all,[160] it is from profit-seeking shareholders' *particular* profit-seeking motives, as opposed to those of the shareholders as a whole.[161]

B. Economic Rights

The classic capital structure of a VC-backed startup includes common stock, held mainly by the startup's entrepreneurs and employees, and convertible preferred stock, held by the startup's investors.[162] Startups that have gone through multiple financing rounds would have several classes of preferred stock outstanding, each held by the respective investors of every such round.[163]

Startups' convertible preferred stock comes with a standard set of favorable economic rights.[164] Importantly, they come with *liquidation*

---

155. *See supra* notes 127–29 and accompanying text.

156. *See supra* notes 92–94 and accompanying text. About the recent changes in OpenAI's organizational structure see *supra* notes 110–111 and accompanying text.

157. For information about OpenAI Global, see *supra* notes 95–98. Similarly, for information about Anthropic, see *supra* notes 133–35.

158. *See supra* note 141 and accompanying text.

159. *Id.*

160. Some have suggested that tie-breaking directors have—or wish to have—long-term relationships with VCs and are therefore not truly independent from them. *See, e.g.*, Fried & Ganor, *supra* note 42, at 988–89.

161. *See* Pollman, *supra* note 11, at 188–200 (discussing archetypal shareholder-level conflicts in VC-backed startups).

162. *Id.* at 170–73.

163. *Id.* at 174.

164. *See* FELD & MENDELSON, *supra* note 11, at 45–105; ROBERT JOE HULL ET AL., REPRESENTING STARTUPS §§ 7:13–7:29 (2023–24 ed.).

preference—the right to be paid, in case of an exit event or distribution of dividends, a certain fixed amount before any proceeds may be paid on common stock or lower-ranking preferred stock.[165] Common shareholders, conversely, may only participate proportionally (pro rata), but without a fixed amount limitation, in the distribution of the remaining proceeds after all preferred shareholder liquidation preferences are satisfied in full.[166] Convertible preferred stock also comes with *conversion rights*, allowing its holders to convert it—actually or notionally—into common stock, forgo its liquidation preference, and participate in the distribution like common shareholders.[167] Naturally, investors will do so if the resulting payment exceeds their liquidation preference. Conversion is the result investors hope for; the fixed liquidation preference is typically designed to act as a safety net, allowing them to recoup no more than their original investment—and perhaps nominal positive returns—in underwhelming exits.[168]

The potential of sharing in the startup's future success is indispensable in creating the right incentives for startup entrepreneurs, investors, and employees to engage and collaborate effectively. Similarly to control rights, entrepreneurs and investors use the allocation of startups' economic rights to bridge information asymmetries and reduce agency costs. For instance, the typical compensation structures of startup entrepreneurs comprise low salaries—compared to the entrepreneurs' labor market value—paired with a significant stake in the startup's equity. This compensation structure is intended to create a strong incentive for entrepreneurs to work diligently to

---

165. Charles Korsmo, *Venture Capital and Preferred Stock*, 78 BROOKLYN L. REV. 1163, 1173 (2014) (defining liquidation preferences); Kristoffer Warren, *Liquidation Preferences*, CARTA (Aug. 7, 2023), https://carta.com/learn/equity/liquidity-events/liquidation-preferences/ [https://perma.cc/4UPW-NPLP].

166. If the liquidation preference is constructed as "participating" preference, the preferred shareholders get to participate again in the distribution of the remaining proceeds. FELD & MENDELSON, *supra* note 11, at 54–63; HULL ET AL., *supra* note 164, at § 7:17.

167. FELD & MENDELSON, *supra* note 11, at 71–77, 88–90; HULL ET AL., *supra* note 164, at §§ 7:18–22. Standard liquidation preference provisions allow preferred shareholders to choose to participate in the distribution on an "as-converted basis" without actually converting to common stock. *Certificate of Incorporation*, NAT'L VENTURE CAP. ASS'NS § 2.1 (Oct. 2025), https://nvca.org/document/nvca-model-certificate-of-incorporation-updated-oct-2025 [https://perma.cc/X3GD-YQLB] [hereinafter *NVCA Model COI*].

168. Non-participating liquidation preference reflecting a 1x multiplier on the original investment amount is a prominent market standard. Peter Walker & Kevin Dowd, *State of Private Markets: Q1 2024*, CARTA (May 7, 2024), https://carta.com/data/state-of-private-markets-q1-2024/ [https://perma.cc/6ECX-WTKV].

increase the startup's enterprise value and minimize the risks of managerial self-dealing or shirking.[169]

The allocation of the startup's economic rights not only allows investors to reduce firm-specific risks by bridging information asymmetries and agency costs; it also enables them to diversify the remaining risk. The VC business model famously relies on "power law" or "long tail" economics.[170] Startups fail at very high rates, and investors count on the returns from a few highly successful startups to compensate for the inevitable failures of most other startups in their portfolios.[171] This mindset is the bedrock of startup finance. The potential of raking in exceedingly high returns from a handful of home runs is more than essential for VC investors; it makes supporting fledgling startups with unproven technologies and inexperienced entrepreneurs into a viable business. Applying the same "power law" logic, even straight debt financings in startups often include an equity component—typically, a warrant—allowing the lender some uncapped stake in the profits of successful startups that would help make up for the expected defaults of less-successful ones.[172]

The uncapped upside potential of startup equity is also crucial in allowing emerging and cash-constrained startups to compete with incumbent firms over top talent. Startups offering equity compensation can attract talent while paying below-market salaries because their equity compensation packages include the potential for significant gains if the startup succeeds.[173]

OpenAI's capped-profit structure was, in this respect, nothing short of revolutionary. While some critics have suggested that the caps placed by OpenAI itself were too high for its equity holders to care about,[174] the concept of limiting investors, entrepreneurs, or equity-holding employees' upside

---

169. Bartlett, *supra* note 152, at 53; Gilson, *supra* note 15, at 1083; Sahlman, *supra* note 15, at 508.

170. Elizabeth Pollman, *Adventure Capital*, 96 S. CAL. L. REV. 1341, 1346 (2024).

171. Elizabeth Pollman, *Startup Failure*, 73 DUKE L.J. 327, 342–43 (2023); Matthew T. Wansley & Samuel N. Weinstein, *Venture Predation*, 48 J. CORP. L. 813, 817–18 (2023).

172. Ibrahim, *supra* note 22, at 1182–84.

173. Yifat Aran, *Beyond Covenants Not to Compete: Equilibrium in High-Tech Startup Labor Markets*, 70 STAN. L. REV. 1235, 1273–76 (2018).

174. *See, e.g.*, Devin Coldewey, *OpenAI Shifts from Nonprofit to 'Capped-profit' to Attract Capital*, TECHCRUNCH (Mar. 11, 2019), https://techcrunch.com/2019/03/11/openai-shifts-from-nonprofit-to-capped-profit-to-attract-capital [https://perma.cc/V46R-QXV6] (discussing how critics have targeted in particular the 100x multiple that applies to OpenAI's early investors); Joyce J. Shen, *Capped Profit in OpenAI*, MEDIUM (Mar. 19, 2023), https://medium.com/@joycejshen/capped-profit-at-openai-f7e67fa93ff3 [https://perma.cc/YXJ3-EZ5G]; David Gray Widder et al., *Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI*, NATURE (forthcoming) (manuscript at 15) (on file with SSRN).

potential was practically unheard of in startups.[175] Furthermore, it appears to defy the basic economic foundations upon which the relationships between startup entrepreneurs, investors, and employees are believed to be built.

C. Choice of Entity

Startups are, by and large, formed as corporations—a practice that legal scholars have observed and extensively discussed over the last few decades.[176] Both OpenAI and Anthropic, however, have made unorthodox decisions regarding their choice of entity.

OpenAI Global was formed as an LLC.[177] LLCs are an uncommon choice for tech startup formation, with their share of US startups estimated to be in the single-digit percentages.[178] Startup entrepreneurs' aversion to forming LLCs is often attributed to their tax treatment.[179] By default, LLCs are pass-through entities whose profits and losses are attributed to their owners, while corporations are taxed separately from their owners.[180] Pass-through taxation may offer significant advantages to owners of Main Street businesses (that are indeed often organized as LLCs).[181] For startups, however, pass-through taxation poses serious difficulties.[182] To name a few, VCs—a primary source of startup financing[183]—are wary about investing in pass-throughs as it might

---

175. The upside potential of common stock held by startup founders and employees, is unlimited by design; they hold the residual claim on the startup's assets. While the preferred stock held by startup investors represents a fixed (or "capped") claim on the startup's assets, the ability to convert their preferred stock into common stock makes investors' claim similarly uncapped. About the typical capital structures of VC-backed startups see *supra* notes 162–68 and accompanying text.

176. *See supra* note 32.

177. *See supra* note 89 and accompanying text.

178. Bo Bian et al., *Conflicting Fiduciary Duties and Fire Sales of VC-Backed Startups* 36, A-23 (June 20, 2022) (on file with Foundations of Law and Finance) (estimating that the average percentage of US startups formed as LLCs in between late 2013 and 2020 was roughly 5.8%).

179. *See generally* Victor Fleischer, *The Rational Exuberance of Structuring Venture Capital Start-ups*, 57 TAX L. REV. 137 (2003) (analyzing why start-up entrepreneurs often do not organize as LLCs); *see also* Polsky, *supra* note 32, at 415 ("[N]ew businesses . . . in Silicon Valley and other hotbeds of VC activities . . . are often told in no uncertain terms to form a corporation."). LLCs may, however, under the IRS' "check-the-box" regime, change their default tax treatment to be taxed separately from their owners. CARTER BISHOP & DANIEL KLEINBERGER, BISHOP & KLEINBERGER ON LIMITED LIABILITY COMPANIES § 2.1 (2025).

180. Polsky, *supra* note 32, at 413.

181. *See id.* at 414–15.

182. *See generally supra* note 179 and accompanying text (discussing startup aversion to organizing as LLCs).

183. For information about VCs as a primary source of startup financing, see Darian M. Ibrahim, *Should Angel-Backed Start-ups Reject Venture Capital?*, 2 MICH. J. PRIVATE EQUITY &

subject their upstream tax-exempt investors to tax reporting in the US.[184] Additionally, equity ownership in pass-through entities is disqualified from consideration as "qualified small business stock" ("QSBS"), a tax status that offers significant benefits for equity holders on exit events[185] (which, according to the classic startup lifecycle, are to be expected early).[186] Pass-throughs might also complicate the tax treatment of employees holding equity incentives—a staple of startup employees' compensation packages.[187]

Non-tax considerations are believed to affect startups' dislike of LLCs as well. For instance, the preferred exit route for startups has traditionally been to go public within a relatively short time from securing their first VC investment.[188] Public markets are presumed to prefer incorporated entities, and formation as an LLC would likely require the startup to convert into a corporation before the initial public offering ("IPO"), incurring transaction costs.[189] A lack of experience with LLCs among startup legal and tax advisors and a less mature legal landscape are also still seen as barriers to their wider adoption by startups, even though LLCs have been around for decades.[190] Lastly, contemporary VC-backed startups' charters are highly standardized and often based on the model charter maintained by the National Venture Capital Association ("NVCA").[191] The NVCA model form was drafted for use by incorporated startups and cannot be used in an LLC setup without significant adaptations.[192]

VENTURE CAP. L. 251, 253–54 (2013); Seth C. Oranburg, *Bridgefunding: Crowdfunding and the Market for Entrepreneurial Finance*, 25 CORNELL J.L. & PUB. POL'Y 397, 404 (2015); Pollman, *supra* note 11, at 170.

184. Fleischer, *supra* note 179, at 158; Polsky, *supra* note 32, at 421–22.

185. I.R.C. § 1202. To qualify for beneficial tax treatment, QSBS must be "stock in a C corporation," which excludes entities classified as a pass-through. *Id.* at § 1202(c)(1); *see generally* Manoj Viswanathan, *The Qualified Small Business Stock Exclusion: How Startup Shareholders Get $10 Million (Or More) Tax-Free*, 120 COLUM. L. REV. F. 29, 38 (2020).

186. According to data gathered by financial data provider PitchBook, between the years 2011 to 2023, the average time from first VC investment to exit in US startups ranged between roughly five to six years for mergers and acquisitions exits and five to eight years for IPO exits. NAT'L VENTURE CAP. ASS'N, 2024 YEARBOOK 32–33 (2024).

187. Fleischer, *supra* note 179, at 169–70; Polsky, *supra* note 32, at 446.

188. *See, e.g.*, Mark A. Lemley & Andrew McCreary, *Exit Strategy*, 101 B.U. L. REV. 1, 17 (2021); Darian M. Ibrahim, *The New Exit in Venture Capital*, 65 VAND. L. REV. 1, 11 (2012).

189. *See* Fleischer, *supra* note 179, at 175; Polsky, *supra* note 32, at 435.

190. *See* Fleischer, *supra* note 179, at 167–73; Polsky, *supra* note 32, at 441–44.

191. Robert P. Bartlett, *Standardization and Innovation in Venture Capital Contracting: Evidence from Startup Company Charters* 5 (Rock Ctr. for Corp. Governance at Stanford Univ., Working Paper No. 253, 2023) (finding, for example, that 85% of startup charters filed in 2022 were based on the NVCA model).

192. *See* NVCA Model COI, *supra* note 167.

Anthropic's decision to organize as a PBC, as well as OpenAI's reclassification of its for-profit entity as a PBC, were more minor deviations from the startup structuring playbook.[193] While LLCs are substantially different from corporations and operate under different legal regimes,[194] PBCs are corporations with a twist. Besides their distinctive rules concerning directors' fiduciary duties,[195] PBCs are subject to general corporate law.[196] Incorporation as a PBC has been observed to be relatively common among technology firms compared to other sectors.[197] Some high-profile examples are the insurance-tech startup Lemonade (which had even gone public as a PBC)[198] and xAI, Elon Musk's AI venture (formed as a Nevada Benefit Corporation).[199] However, in absolute numbers, PBCs remain an unusual choice for startup formation.[200] A 2021 study by Dorff et al. of Delaware-registered PBCs found that between 2013 and 2019, only 707 investments—most of which by VCs and other typical startup investors—were made into PBCs, compared to 50,000 investments made by VCs in the same period overall.[201]

Various reasons have been offered to rationalize PBCs' *relative* popularity among technology firms.[202] These include, for instance, the industry's typical vying for younger customers and employees,[203] its disproportionally significant stake in forestalling unfavorable regulation through self-regulation,[204] and its increased reliance on long-term business models that are, arguably, more easily defendable from fiduciary challenges under a PBC.[205] The reason PBCs have yet to become more popular *per se* has been mostly

193. *See supra* notes 122–27 and accompanying text.

194. In Delaware, corporations are governed by DEL. CODE ANN. tit. 8, chapter 1, and LLCs by DEL. CODE ANN. tit. 6, chapter 18.

195. *See infra* notes 295–96 and accompanying text.

196. DEL. CODE ANN. tit. 8, § 361 (West 2013).

197. They are similarly popular among the consumer products and services sector. *See* Michael B. Dorff et al., *The Future or Fancy? An Empirical Study of Public Benefit Corporations*, 11 HARV. BUS. L. REV. 113, 143–45 (2021).

198. Robert S. Rachofsky et al., *Lemonade, Inc.: Harbinger of Future Public Benefit Corporation IPOs?*, CLS BLUE SKY BLOG (Aug. 14, 2020), https://clsbluesky.law.columbia.edu/2020/08/14/lemonade-inc-harbinger-of-future-public-benefit-corporation-ipos [https://perma.cc/DX6U-5N34].

199. Peterson, *supra* note 53.

200. *See* Dorff et al., *supra* note 197, at 116–17.

201. *Id.* at 116–17, 145, 151.

202. For discussion and potential rebuttals, see generally Alanna Potter, *Purpose or Profit? The Rise of Public Benefit Corporations in the Technology Industry*, 20 DUKE L. & TECH. REV. 90 (2023) (analyzing reasons why technology organizations have adopted PBCs).

203. *Id.* at 94–104.

204. *Id.* at 104–09.

205. *Id.* at 109–14.

associated with the legal uncertainty revolving around them.[206] Since PBCs are relatively new and the lack of cases interpreting their unique fiduciary regime, PBC directors cannot be sure about the exact scope of protection they enjoy when taking stakeholders' interests into account.[207] Outside the VC-funded world, similar arguments are often brought up in the responses of publicly traded firms to shareholder proposals seeking to have them reclassified as a PBC.[208]

III. THE USE CASE: ALIGNING STARTUPS' AND HUMANITY'S INCENTIVES

The previous section described how OpenAI and Anthropic departed from conventional startup structuring. A central motivation for the unique features of their business structuring models, as discussed briefly above and in more detail below, is the hope of cultivating a more robust commitment to AI safety and ethics among decision-makers and equity holders.

While these experiments may be idiosyncratic in design, the governance challenges they aim to address are not necessarily so—and presumably apply to other AI startups as well. One set of challenges involves limiting the influence of profit-maximizing directors or managers who might disregard public safety or ethical risks ("the *greedy director*" problem). Another involves shielding public-minded ("*prudent*") directors or managers from interference by equity holders who prioritize returns above all else ("the *greedy shareholder*" problem).

At this point, the Article shifts focus. Rather than evaluating OpenAI or Anthropic's specific experiences, it asks a broader and more policy-relevant question: are similar models likely to help protect society from the risks posed by advanced AI—and should policymakers require or encourage their adoption? Should Alignment Stewards, profit-capping mechanisms, and alternative entity forms be treated as governance features that every AI startup ought to implement? Should government officials, as the drafters of the *Not for Private Gain* letters had suggested,[209] be concerned about an AI startup rolling back an Aligned Structuring model (or relieved to see one being adopted)?

---

206. Dorff et al., *supra* note 197, at 152–55.

207. *Id.*

208. *See, e.g.*, Jill E. Fisch, *Purpose Proposals*, 1 U. CHI. BUS. L. REV 113, 128–29 (2022) (discussing such proposal brought by Wells Fargo shareholders); RICHARDS, LAYTON & FINGER, REPORT TO THE BOARD OF DIRECTORS OF JPMORGAN CHASE & CO. REGARDING PUBLIC BENEFIT CORPORATIONS 6–7 (2021).

209. *See supra* notes 105–06 and accompanying text.

While OpenAI and Anthropic's efforts offer useful case studies, this paper does not attempt to judge the success or failure of their structures or generalize from them directly. A sample of two firms is too small to isolate firm-specific effects. Instead, the goal is to evaluate the theoretical logic of Aligned Structuring as a concept: whether, in principle, its main design elements could meaningfully improve alignment with societal interests, and whether they are scalable in a way that justifies broad policy endorsement.

To justify broad policy endorsement, two conditions must be met. *First*, the model must work and scale. They must be able to meaningfully influence the behavior of startup decision-makers so they internalize safety and ethical concerns they might otherwise neglect, and do so across a range of AI startups, not just rare outliers. If Aligned Structuring's effectiveness depends on unusual traits (which, perhaps, only apply to OpenAI, Anthropic, and a handful of other startups), it cannot serve as a general solution. *Second*, the model must be efficient: the costs of adopting it must not outweigh the societal benefits of improved interest alignment. Startups, as discussed above, are typically structured to maximize their access to capital and talent.[210] If Aligned Structuring burdens or undermines these goals too severely, and the resulting slowdown in innovation produces net harm to society, then even a well-intentioned model may have a net-negative impact on society's interests.

The rest of this section critically examines the three distinctive features of OpenAI and Anthropic's models—Alignment Stewards, profit-capping, and alternative entity forms—through this lens. It considers how each mechanism is supposed to improve alignment, how sound the theory behind it is, and how feasible or costly it would be for the average AI startup to incorporate.

A. Aligned Control Structures

Startups inspired by OpenAI and Anthropic could incorporate Alignment Stewards into their control structures in various ways. The extent of control granted to them can range from complete control, as in OpenAI's model,[211] to blocking power on specific matters, as in the initial stage of Anthropic's model,[212] to a purely advisory role. While OpenAI and Anthropic have chosen to build their own Alignment Stewards from scratch,[213] other options exist. Given sufficient demand, the market might witness the emergence of professional Alignment Stewards that would hold control rights in multiple

---

210. *See supra* Part II.
211. *See supra* notes 92–94 and accompanying text.
212. *See supra* notes 127–29 and accompanying text.
213. *See* discussion *supra* Section I.B.

startups or advise their boards on safety and ethics without holding explicit control rights. Professional Alignment Stewards could leverage economies of scale and scope, making their services more affordable for fledgling startups than maintaining their own firm-specific Alignment Steward.

Combining Alignment Stewards in a startup's control structure, regardless of the method, must meet two requirements to be effective. *First*, the control rights transferred to the Alignment Steward on paper should also provide the desired extent of *de facto* control. Otherwise, the Alignment Steward may not have enough influence to foster a greater commitment to safety and ethics. *Second*, the Alignment Steward must be a better decision-maker—meaning, less likely to cause social harm—compared to the startup's financially motivated stakeholders when considering safety and ethics issues. Otherwise, even if they have the desired extent of influence, the Alignment Stewards would not make the startup's conduct more safe or ethical. The rest of this subsection will examine how likely Aligned Structuring models are to satisfy each requirement.

1. Blindspot No. 1: Informal Control in Startup Governance

Aligned control structures aim to foster a stronger commitment to safety and ethics by redistributing the startup's formal controls, such as the legal authority to dictate safety and ethics policies or, at least, appoint officers with the power to do so. However, the allocation of formal control is just one component of a startup's governance structure and may not always be the most important one. Startup stakeholders rely on a complex array of incentives, sanctions, and informal norms to guide the startup's behavior, which can often override the impact and potency of formally assigned control rights.

VCs, for instance, have been long observed to stage their investments strategically as a means of control. Rather than providing startups with the total amount of anticipated funding upfront, VCs spread their investments across multiple rounds throughout the startup's life.[214] Startups, especially at their earlier stages, are typically cash-constrained and heavily reliant on investors' capital to sustain their ambitious growth objectives. As a result, VCs possess an increasing leverage over their portfolio startups' management teams as the startups gradually burn through their cash reserves.[215] While a

214. *See* Gilson, *supra* note 15, at 1073.

215. For information about the use of staged financing by VCs, see Gilson, *supra* note 15, at 1078–81; Pollman, *supra* note 11, at 186–88; Sahlman, *supra* note 15, at 506–08; Smith, *supra* note 11, at 323–24.

startup may approach other VCs or alternative funding sources if its existing VC investors refuse to provide further funding, the existing VCs' hesitation might send a negative market signal, making it significantly more difficult for the startup to secure outside funding on reasonable terms.[216] Staged financing is considered one of the most potent means of control in startup investors' toolboxes, particularly in the early stages. William Sahlman, in a landmark paper, described it as "the key to the relationship between the entrepreneur and the venture capitalist."[217] A VC may use the leverage earned by staged financing to require that a startup changes its business strategy to its liking as a condition for providing further funding.[218] In the safety and ethics context, a greedy VC may demand, for example, that a startup expedite the release of potentially unsafe or unethical products as a condition for providing further funding.

VCs are not the only ones wielding informal control over startups. Big Tech firms, which are major investors in the AI startup scene,[219] also hold significant informal influence. Their power stems from their unique ability to offer both the financial capital and the immense computing power and training data that AI startups crave, especially those developing foundation models.[220] The collaboration between OpenAI and Microsoft, where Microsoft supplies both capital and computing resources, has been criticized for giving Microsoft significant leverage, if not *de facto* control, over OpenAI's management.[221] Big Tech and other corporate investors can leverage this influence to steer AI startups' business strategies towards their own investment goals—like early access to the startup's technology or boosting cloud computing sales[222]—which could potentially come at the expense of safety or ethical considerations.

The use of informal control is not limited to the investor side of startups' governance structures; entrepreneurs, too, may hold a significant extent of informal control. Their leverage can stem from their singular expertise, strong relationships with stakeholders, or reputation that helps attract capital, talent,

---

216. Gilson, *supra* note 15, at 1079–80; Sahlman, *supra* note 15, at 507.

217. Sahlman, *supra* note 15, at 507; *see also* Bartlett, *supra* note 152, at 54; Jason M. Gordon & David Orozco, *Trust and Control: The Value Effect of Venture Capital Term Sheet Provisions as Risk Allocation Tools*, 4 MICH. BUS. & ENTREPRENEURIAL L. REV. 195, 217 (2015).

218. *See* Smith, *supra* note 11, at 324 (referring to VCs' use of the leveraged earned through staged financing when negotiating over the startup's exit strategy).

219. *See infra* note 268 and accompanying text.

220. *See supra* note 89 and accompanying text.

221. Kak et al., *supra* note 16; Complaint at ¶ 125, Musk v. Altman, (Cal. Super. Ct. filed Feb. 29, 2024) (No. CGC-24-612746), 2024 WL 899024; Pedersen, *supra* note 57, at 31.

222. *See infra* notes 265–66 and accompanying text.

and business—making them "Superstar CEOs" whose effective control may extend far beyond their formal control rights.[223] VCs have been known to make their investment decisions, particularly in the early stages, based on the strength of the entrepreneurial team, and an entrepreneur's premature departure could leave investors with a stake in a firm that has essentially become, at least from their perspective, an empty shell.[224] A valuable entrepreneur's threats to abandon a startup due to disagreements over its policy on safety and ethics could risk the startup's future no less than a VC's threat to withhold further funding.

The failed attempt to oust OpenAI's CEO, Sam Altman, in late 2023 has often been described as an example of how stakeholder influence can render formal control structures in AI startups nearly meaningless.[225] On November 17, OpenAI's board announced Altman's removal as CEO and board member.[226] In the following days, OpenAI's board has reportedly encountered immense pressure from OpenAI's investors and employees to reinstate Altman as CEO, which included Microsoft announcing that Altman would be joining its ranks to lead a new AI team; almost all of OpenAI's employees signing a letter threatening to quit if Altman was not reinstated; and investors considering writing off the value of their holdings in OpenAI to zero.[227] Just five days later, the board reversed course, reinstating Altman as CEO and dramatically reshaping the board, with three out of four pro-removal directors being removed and replaced, with some of the replacements alleged to have been "hand-picked" by Altman and "blessed" by Microsoft.[228]

Even with optimal Alignment Stewards in formal control, startups might still succumb to greedy stakeholders' informal influence—at least to some extent. Optimal Alignment Stewards, that are perfectly aligned with society's interest, will aspire to maximize the "net" social value of new AI products—doing what they can to increase their expected social benefits and decrease

---

223. *See* Hamdani & Kastiel, *supra* note 17, at 1366–96; Tornetta v. Musk, 310 A.3d 430, 507–08 (Del. Ch. 2024).

224. Nikolaus Franke et al., *Venture Capitalists' Evaluations of Start-Up Teams: Trade-Offs, Knock-Out Criteria, and the Impact of VC Experience*, 32 ENTREPRENEURSHIP THEORY & PRAC. 461–62 (2008); Paul A. Gompers et al., *How Do Venture Capitalists Make Decisions?* 135 J. FIN. ECONS. 169, 178 (2020).

225. Andhov, *supra* note 56, at 9–10; Kak et al., *supra* note 16; Matt Levine, *Who Controls OpenAI?*, BLOOMBERG (Nov. 20, 2023), https://bloom.bg/3LLFikS [https://perma.cc/ZP4Q-D4SD]; Pedersen, *supra* note 57, at 31–32.

226. Pedersen, *supra* note 57, at 21–22.

227. *Id.* at 22–24.

228. *Id.* at 24–26; Complaint, Musk v. Altman, *supra* note 221, at ¶ 31 ("On information and belief, the new Board members were hand-picked by Mr. Altman and blessed by Microsoft.").

their potential risks. While they will never release a net-negative product, they may tolerate *some* level of risk provided that the product's societal benefits are significant enough. When having to choose between releasing a net-positive product with a less-than-ideal risk profile or no product at all, they would choose the former. This fact allows profit-seeking stakeholders some leverage, contingent upon the startup's dependency on their unique contributions.

Assume, for instance, that an early-stage startup controlled by an optimal Alignment Steward is working on its first product, which would have a net social value of V if released. By developing a certain safety feature and incorporating it into the product, the startup can increase the product's net social value by 10, while making it somewhat less profitable. Assume next that the startup is severely cash-constrained and financially dependent upon its profit-seeking investor. The investor, indifferent to the product's societal value, threatens to withhold all further funding and force the startup to shut down if the product is not released immediately. Shutting down before product release would have a V value of 0. Figure 1 below displays the Alignment Stewards' expected responses to the VC pressure as a function of V.

Figure 1. Alignment Steward Responses to VC Pressure as a Function of Net Social Value (*V*)

V Value Without (With) Safety Feature	Develop Feature?	Release Product?	Social Impact of VC Pressure
-15 (-5)	No	No	0
-5 (5)	No	No	-5
5 (15)	No	Yes	-10

With the *V* value set at -15, the VC pressure makes no difference. With VC pressure or without it, the Alignment Steward would not have developed the safety feature nor released the product since its *V* value would nevertheless remain net-negative. It would rather see the startup go bust (*V=0*). With a *V* value of -5, however, the VC's pressure has a negative social effect. The Alignment Steward would rather go bust (*V=0*) than release an unsafe product (*V*= -5). However, free from VC pressure, it would have developed the safety feature and released a net-positive product (*V*=5). With an initial *V* set at 5, the VC's impact is even more pronounced. Free from VC pressure, the Alignment Steward would have developed the feature and released a product with *V* value of 15. However, it would prefer to succumb to the VC's pressure and release the product without the safety feature (*V*=5) than go bust (*V*=0).

Caving in to stakeholder pressure, therefore, does not necessarily indicate corruption or insincerity on the part of the startup's management. It could also reflect the limits of their bargaining power and genuine efforts to make rational decisions that best serve society's interests. To completely insulate managers from profit-seeking stakeholder pressure, it is not enough to ensure their interests align with society's. The managers in control must be "fundamentalists" of AI safety and ethics, determined to minimize risks at all costs—which, in fact, would have adverse implications on society's interests.

The *Not for Private Gain* letter drafters placed considerable weight on the importance of allocating formal control within OpenAI – particularly on the idea that "full management control" retained with the nonprofit parent, including the power to fire executives and directors in the for-profit entity, is essential to maintaining alignment with safety and public interest goals.[229] That emphasis is not entirely misplaced; allocating formal controls is undeniably an important aspect of designing any startup's governance structure.[230] But the analysis above shows how focusing on formal controls

229. *See Not for Private Gain Second Letter*, *supra* note 109, pts. II.2, III.2(a).

230. *See supra* notes 138–43 and accompanying text.

in isolation risks missing the broader reality. An AI startup's operating environment requires controllers who are committed to society's interests to practice a fair degree of realpolitik in their dealings with profit-driven stakeholders and reach compromises where necessary. Assigning formal control to an Alignment Steward, therefore, does not mean giving them complete control over *all* strategic decisions in any meaningful sense, just as shifting formal control away from them does not necessarily afford financially motivated stakeholders an extent of control that they did not already hold. Informal levers of influence—cash, computing power, relationship with key personnel, entrepreneurial star power—can be no less potent than board seats and voting rights. Any serious analysis of control in AI startups must therefore go beyond the allocation of formal controls and ask: who, of all the startup's stakeholders, is most indispensable to the startup's survival and success?

2. Blindspot No. 2: Alignment Stewards' Competence

The managers of AI startups do not act as agents of society as a whole in a strict legal sense. However, agency cost analysis, often used to analyze the relationship between a firm's shareholders and managers, helps illustrate the root problems for which AI startups might put society's interests at risk and the limits of Aligned Structuring in addressing them.[231]

When startup managements' decisions optimally serve societal interests—releasing all net-positive products and technologies while maximizing their net social value and never releasing net-negative ones—they function as perfect agents, imposing no avoidable AI risks on society. However, two archetypal governance problems might cause AI startup managers to act as imperfect agents, imposing avoidable costs and risks on society.

The first is a *conflict problem.* If a product or technology yields a net negative for society but a net positive for the startup's managers, rational managers will likely release it. This can occur when managers reap benefits not shared by society (for example, performance-based compensation tied to sales milestones) or are somehow less affected by the product's safety or ethics risks (for example, are able to avoid using it personally). The additional AI risks to which society may be exposed due to this divergence of interests can be termed "conflict costs."[232]

---

231. *See also* Tallarita, *supra* note 19; Ofer Eldar, *Designing Business Forms to Pursue Social Goals*, 106 VA. L. REV. 937, 941–43 (2020) (referring to the governance of prosocial businesses generally).

232. For information about conflict costs, see Goshen & Squire, *supra* note 20, at 790–95.

The second governance problem is a *competence problem*. Startup managers may genuinely strive to protect society's interests but fail to accurately assess the net social value of new technologies or products. They might release harmful products by underestimating risks or overestimating benefits or deny society safe and valuable products by overestimating risks or underestimating benefits. The additional costs to which society may be exposed due to these unintentional oversights can be termed "competence costs."[233]

The less conflicted and more competent startup managers are—or, put differently, the lower the combined conflict and competence costs they impose on society—the better decision-makers they are from society's perspective.[234] The use of Alignment Stewards is designed to minimize managers' conflict costs by placing control in the hands of individuals who presumably have no personal stake in the startup's financial performance—a potentially significant source of misalignment between managerial and societal interests. To achieve this, Aligned Structuring models may stipulate that Alignment Stewards have no personal financial interest in the startup,[235] are appointed and removed by peers rather than profit-seeking stakeholders,[236] or serve predetermined, short tenures.[237] The *Not for Private Gain* letter drafters echoed the same logic, emphasizing whether, after its then-proposed restructuring, OpenAI would retain a majority-independent board[238] or remain subject to Attorney General oversight (as nonprofits are, but PBCs are not).[239] These are all variations on the same theme: efforts to ensure the Alignment Steward is indeed properly aligned with society's interest, which in turn would reduce conflict costs between society and the AI developers, who act as its notional agents.[240]

The effectiveness of each of these alignment measures may vary greatly, and none should be considered bulletproof.[241] However, it would be hardly

---

233. For information about competence costs, see *id.* at 785–90.

234. *Id.* at 783–85.

235. *See Capped Profit Structure Primer*, *supra* note 13 and accompanying text; *LTBT Primer*, *supra* note 6 and accompanying text.

236. *See Capped Profit Structure Primer*, *supra* note 13 and accompanying text; *LTBT Primer*, *supra* note 6 and accompanying text.

237. *See LTBT Primer*, *supra* note 6 (Anthropic).

238. *Not for Private Gain First Letter*, *supra* note 1, pt. II.A.4.

239. *Id.* pts. II.A.2, IV.2.3.

240. *See* Eldar, *supra* note 231 and accompanying text.

241. *See, e.g.*, Ofer Eldar & Mark Ørberg, *The Anatomy of Nonprofit Control of Business Enterprise*, 39 YALE J. REG. (forthcoming 2026) (suggesting that attorneys general "theoretically oversee nonprofits under state charitable trust laws, but enforcement is often weak, with limited resources and infrequent regulatory actions"); Tallarita, *supra* note 19 (arguing that "[s]ocially-

controversial to say that they make Alignment Stewards at least *somewhat less* conflicted than downright profit-seeking managers—and undoubtedly, society is better off when those controlling powerful AI systems have fewer incentives to act against its interests and conflict costs are reduced to a minimum. But even assuming a perfect Alignment Steward—one entirely insulated from pecuniary interests and subject to effective societal oversight and enforcement powers (by an Attorney General or otherwise)—eliminating all conflict costs can only get society part of the way there. Equal attention must be paid to the competence costs of a given governance structure. If Alignment Stewards cannot identify AI risks when they arise or determine appropriate responses at least as well as profit-oriented stakeholders can, their good intentions alone will not make them better decision-makers, all things considered.

At a high level of abstraction, competence in identifying safety and ethics risks comprises two distinct skill sets. The first is "technical" skills, referring to the director's understanding of the startup's technology (e.g., its operation and potential for uncontrolled behavior). The second is "policy" skills, referring to the director's grasp of the technology's potential societal implications (e.g., legal, reputational, ethical, or x-risk consequences of different AI capabilities and behaviors).[242] These skill sets vary in relevance to addressing specific safety or ethics risks. Addressing some risks primarily requires technical skills, while addressing others demands strong policy skills. A director's overall competence increases the more skilled they are, and the more relevant their skills are.

Startups using Alignment Stewards can enhance their levels of skill by appointing qualified individuals with diverse backgrounds as directors. Anthropic's LTBT statement, for instance, highlights how LTBT trustees possess "backgrounds and expertise in AI safety, national security, public policy, and social enterprise,"[243] and OpenAI's board, as of February 2026,

---

oriented governance structures should not be content with independence from executives and investors" and noting that "[d]irectors who cannot be fired by investors are less likely to follow investor preferences, but are they more likely to choose what is best for society?").

242. *See, e.g.*, Kevin Buehler et al., *Getting To Know—and Manage—Your Biggest AI Risks*, MCKINSEY ANALYTICS (May 3, 2021), https://www.mckinsey.com/capabilities/quantumblack/our-insights/getting-to-know-and-manage-your-biggest-ai-risks [https://perma.cc/8VMU-PR3N] (noting that "organizations must put business-minded legal and risk-management teams alongside the data-science team at the center of the AI development process" to protect against AI risks effectively); NAT'L A.I. ADVISORY COMM., FINDINGS & RECOMMENDATIONS: AI SAFETY 5 (2024) ("Safety standards will benefit from multidisciplinary expertise, ranging from technical expertise needed to audit complex systems to social science and humanities expertise to assess the broader contexts around AI deployments.").

243. *LTBT Primer*, *supra* note 6.

was populated by notable figures in similar fields.[244] Nevertheless, even with highly skilled Alignment Stewards, it is improbable that they can match the technical skill of a specific type of financially interested directors: the startup's entrepreneurs.

As the technological visionaries behind the startup, entrepreneur-managers are presumed in startup scholarship to possess an extreme information advantage over non-entrepreneur managers (which, in startup governance, typically means those appointed by investors)[245] in all matters relating to the startup's technology.[246] The entrepreneurs' information advantage increases the more groundbreaking or technologically complex the technology or product is.[247] It is also expected to be particularly emphasized at the startup's earlier, pre-product stage, which focuses on reducing the entrepreneurs' abstract ideas into lines of code.[248] At this embryonic phase, it would be impossible for outsiders to independently observe or verify any information about the future technology or product's risk profile.

Startup investors, just like Anthropic and OpenAI's Alignment Stewards, often try to bridge competence gaps by appointing team members with relevant skills and experience (e.g., tech-savvy individuals, ideally with prior entrepreneurial experience) or by specializing in narrow industry domains.[249] However, even with extensive and relevant technical knowledge, entrepreneurial background, and deep knowledge of the relevant industry, non-entrepreneur directors would rarely catch up with the entrepreneurs' information advantage on technology-related matters.[250] Directors are, by definition, high-level decisionmakers. Unless they also hold other managerial

---

244. As of February 2026, the board of OpenAI's nonprofit parent included Sam Altman; Adam D'Angelo, CEO of Quora; Sue Desmond-Hellman, ex-CEO of the Bill & Melinda Gates Foundation; Paul M. Nakasone, a retired U.S. Army General; Nicole Seligman, ex EVP and General Counsel at Sony; Bret Taylor, ex co-CEO of Salesforce; Zico Kolter, a Carnegie Mellon machine learning professor; and Adebayo Ogunlesi, managing partner at the private equity fund GIP. *Our Structure*, *supra* note 97.

245. *See supra* notes 138–40 and accompanying text.

246. Gilson, *supra* note 15, at 1077; Ibrahim, *supra* note 22, at 1192; Pollman, *supra* note 22, at 628.

247. Gilson, *supra* note 15, at 1077; Ibrahim, *supra* note 22, at 1192; Pollman, *supra* note 22, at 628.

248. Gad Weiss, *A Theory of Seed Financing*, 2025 U. ILL. L. REV. 1831, 1873–76 (2025).

249. Guarav Jain, *How to Be a VC Without Any Capital*, TECHCRUNCH (Sept. 19, 2015), https://techcrunch.com/2015/09/19/how-to-be-a-vc-without-any-capital/ [https://perma.cc/N6DZ-CMZB] (referring to VCs); Darian M. Ibrahim, *The (Not So) Puzzling Behavior of Angel Investors*, 61 VAND. L. REV. 1405, 1431–32 (2008) (referring to angel investors).

250. Gilson, *supra* note 15, at 1077.

positions (which the entrepreneurs, unlike investors, often do),[251] they will not be getting their hands dirty in the startup's day-to-day activities.[252] Even if they possess the exact extent of technical knowledge and experience as the entrepreneurs, they will certainly not have invested the same amount of human capital as the entrepreneurs, working for months or years on the particular problem the startup aims to solve.

Having the entrepreneurs themselves on the board or advising the controlling Alignment Steward on matters requiring technical skill can help non-entrepreneur directors make better decisions on technology-related matters but will not completely alleviate the effect of the entrepreneurs' relative advantage. Assume, for instance, that the entrepreneurs correctly advise the Alignment Steward that a certain product feature raises no significant safety or ethics concerns. Alignment Stewards with inferior technical skills may overrule them and prevent the product's release anyway, erroneously suspecting that the entrepreneurs' financial or other personal interests cloud their judgment.[253] The more significant the difference between the entrepreneurs' and the Alignment Stewards' technical skills, and the less the Alignment Stewards trust the entrepreneurs' intentions, the more likely the Alignment Stewards would be to make such false-positive calls.[254]

While Alignment Stewards can take measures to increase their levels of skill, the *relevance* of their skill sets depends on the startup's technology or product—not the directors' particular traits. Some technologies or products are inherently complex to understand or control, while their potential social implications are clear and straightforward, requiring no specialized skills to recognize. For example, consider an AI-powered nuclear missile control

---

251. Importantly, an entrepreneur usually acts as the startup's CEO, at least at its earlier stages. *See* Brian Broughman & Jesse M. Fried, *Do Founders Control Start-up Firms that Go Public?*, 10 HARV. BUS. L. REV. 49, 55–56 (2020).

252. Bret Taylor, chair of OpenAI's board, has been quoted saying that his "job at OpenAI is really not around day-to-day operations but governance." Kylie Robinson, *OpenAI Chair Bret Taylor Says He'll Recuse Himself 'Whenever There Is a Potential for Overlap' With His New AI Startup Sierra*, YAHOO! FINANCE (Feb. 13, 2024), https://yhoo.it/4cfbCaz [https://perma.cc/6JAS-U5SD]. *See generally* Andrew A. Schwartz, *The Perpetual Corporation*, 80 GEO. WASH. L. REV. 764, 770 (2012).

253. This scenario is an agency problem within an agency problem, where the entrepreneurs act as notional agents of the Alignment Stewards (which themselves act as society's notional agents). The negative impact of the Alignment Stewards' wrongly overruling the entrepreneurs' advice may be termed "principal competence costs." Goshen & Squire, *supra* note 20, at 786–88.

254. *See id.* at 787. Having the entrepreneurs advise on matters requiring technical skill is more likely to assist in preventing *false-negative* calls by Alignment Stewards. If profit-seeking entrepreneurs with superior technical skill advise that a certain product is net-negative, the Alignment Stewards would have no apparent reason to question their judgment.

system with a black-box target selection algorithm.[255] While assessing the system's likelihood of defying human control can be technically challenging, the catastrophic social consequences of such loss of control are easily foreseeable. In this case, directors' ability to prevent these risks would primarily depend on their technical skills, while differences in their policy skills would be negligible.

Conversely, imagine an AI that creates personalized music playlists based on the user's listening history and demographics. While the algorithm may be simple to understand by directors with high and low technical skills alike, it could raise complex and nonobvious societal concerns. For instance, could the algorithm, by limiting user exposure to unfamiliar artists and genres, reinforce existing cultural biases beyond musical tastes?[256] Or, how likely is it that such an algorithm undermines cultural diversity?[257] In this scenario, directors' technical skill levels become negligible when assessing their competence to mitigate risks. Instead, their relative policy skills become the most important factor.

Transferring control to Alignment Stewards thus does not guarantee that control would be placed with the agent most likely to make the right decisions to protect society's interests. While Alignment Stewards directors' conflict costs are likely to be lower than those of profit-seeking directors, a competence costs analysis provides a more complex picture. Of the three determinants of directors' overall competence in identifying safety and ethics risks—technical skills, policy skills, and the relevance of these skill sets—one crucial factor characteristically favors entrepreneur-directors, and another depends on the specific traits of the startup's technology or product, and the nature of potential safety and ethics risks it may create. Given that the ideal decision-maker minimizes conflict *and* competence costs, Alignment Stewards' lower conflict costs—even if coupled with high policy skill—would not inherently make them superior to profit-seeking ones. In startups where technology necessitates high technical skills to mitigate risks, profit-seeking entrepreneurs might be better decision-makers despite their profit-seeking motivations.

---

255. A similar AI was famously portrayed in the science fiction classic Wargames. WARGAMES (United Artists & Sherwood Productions 1983).

256. For a discussion about the potential connection between musical preferences and listeners' attitudes to different cultures, see Eric Clarke et al., *Music, Empathy and Cultural Understanding*, 15 PHYSICS LIFE REV. 61 (2015); Heather L. LaMarre et al., *Does the Music Matter? Examining Differential Effects of Music Genre on Support for Ethnic Groups*, 56 J. BROAD. & ELEC. MEDIA 150 (2011).

257. *See, e.g.*, Michal Shur-Ofry, *Multiplicity as an AI Governance Principle*, 100 IND. L. J. 1527 (2025).

B. Aligned Capital Structures

OpenAI's announcements indicated that capping equity holders' profits, with the excess going back to the nonprofit parent, served two distinct purposes. The first was "to ensure that most of the value (monetary or otherwise) we [at OpenAI] create if successful benefits everyone"[258]—a distributional purpose which is a vestige of OpenAI's origins as a pure nonprofit.[259] The second purpose, however, went beyond simple distribution of wealth, seeking "to incentivize [investors and employees] to research, develop, and deploy AGI in a way that balances commerciality with safety and sustainability, rather than focusing on pure profit-maximization."[260] That is, profit-capping, much like assigning control to Alignment Stewards, appears to have also been harnessed to help align equity holders' interests with society's broader interests.

The logic behind using profit-capping for interest alignment purposes seems to be the following. Investors and employees presumably have some given extent of prosocial tendencies—they are willing to sacrifice some personal gains to protect society's broader interests, but only up to a point, and assuming that the potential benefits for society are worth it.[261] If their potential financial profits become high enough, or the benefits for society are too minor, the investors and employees might be tempted to chase them while disregarding safety or ethics concerns. For instance, they may oppose managerial decisions that would have a positive societal effect but a negative effect on their holdings' value.

For this theory to work, two conditions must be met. *First*, the profit-capping mechanism used must influence equity holders' internal calculations. The equity holders in question must be sensitive enough to changes in their profit-making potential, and the profit-capping mechanism must not be easy to circumvent or manipulate. *Second*, capping profits must not create unintended consequences that counteract the desired effect of making equity holders less likely to prioritize profits over safety and ethics. As discussed below, it is often unclear whether profit capping would meet either requirement at a given startup.

---

258. *Capped Profit Announcement*, *supra* note 24.

259. See *supra* notes 82–84 and accompanying text.

260. *Capped Profit Structure Primer*, *supra* note 13.

261. For a discussion of the term "prosocial," see LYNN STOUT, THE SHAREHOLDER VALUE MYTH 9 (2012); Patrick M. Corrigan, *The Corporate Governance Trilemma*, 17 J. LEGAL ANALYSIS 141, 146–47 (2025).

1. Blindspot No. 1: Alternative Profit Channels

While profit-capping targets equity holders' potential to make direct financial returns from investing in startups, some startup equity holders rely on factors other than their express economic rights to make positive return-on-investment. *Corporate investors* are the most prominent example. Big tech firms—particularly tech giants like Microsoft, Google, Nvidia, and Amazon—are incredibly active investors in AI startups, either directly from their balance sheets or through their affiliated corporate venture capital arms (CVCs).[262] In 2023, deals involving corporate investors provided a striking 89.7% of all capital invested in generative AI startups worldwide.[263] As mentioned above, the significance of corporate investors for AI startups goes beyond financial support; they are also a primary source for the enormous computing power and training data that AI startups need to develop their technologies and products.[264]

Corporate investors' investment thesis differs from that of traditional startup investors. Rather than focusing exclusively on financial returns, corporate investors are often also concerned about their investments' "strategic" aspects.[265] These may include the potential for gaining early access to the startup's technology and talent, supporting the rise of new markets that could drive demand for the investors' products, or dipping their toes in new markets before deciding to enter them more decisively.[266] The strategic aspects of corporate investments are often the most prominent ones, with financial returns ranking anything between a second-grade consideration and a "nice-to-have"; Strebulaev and Ying Wang's survey of seventy-nine CVCs revealed that only 11% of the CVCs interviewed have reported placing higher priority on financial rather than strategic investment objectives.[267]

Microsoft and OpenAI's business collaboration is a prominent example of corporate investors' possible strategic objectives when investing in AI

---

262. *Generative AI Bible*, *supra* note 27, at 52; PITCHBOOK GENERATIVE AI REPORT, *supra* note 27, at 8–9. For a discussion about CVCs, see generally Alon-Beck, *supra* note 27; Chesbrough, *supra* note 27; Fan, *supra* note 27.

263. PITCHBOOK GENERATIVE AI REPORT, *supra* note 27, at 8 (based on data compiled as of Oct. 15, 2023).

264. *See supra* notes 218–21 and accompanying text.

265. Alon-Beck, *supra* note 27, at 1018–19; Fan, *supra* note 27, at 369.

266. *See* Chesbrough, *supra* note 27 for a detailed review of CVC strategic investment thesis.

267. Strebulaev & Ying Wang, *supra* note 28, at 23.

startups.[268] The new trend of "Reverse Acquihires", or quasi-acquisitions of AI startups by Big Tech that emerged in early 2024 also helps illustrate what corporate investors' ex-ante investment objectives in AI startups might be.[269] In this new "bailout" deal structure, instead of acquiring a struggling startup as a legal entity, a corporate acquirer hires the employees of the startup (in which it may have previously invested) and licenses its technology, with the licensing fees being used to pay back the startup's investors.[270] At least three high-profile AI startups—Adept AI, Character.AI, and Inflection—have been reported to have struck deals of similar nature with Amazon, Google, and Microsoft, respectively.[271]

Related-party transactions are another profit channel that equity holders may have on top of their formal economic rights. OpenAI, for instance, was reported to do business, or to have been in talks about doing so, with various companies in which Sam Altman was said to hold interests at the time—including Helion, a fusion energy startup in which Altman then served as chairman, and the forum platform Reddit, in which Altman was then a minority shareholder.[272] Where a startup equity holder also transacts with the startup, directly or through affiliated parties, its financial stake in the startup's economic performance could become significantly greater than what its express economic rights imply.[273]

Related-party transactions usually raise corporate governance concerns when reflecting above-market terms for the firm manager or shareholder on both sides.[274] However, even a related-party transaction that is done strictly on market terms—or even on terms favorable for the startup—can reduce a startup manager's or equity holder's reliance on its express, and potentially capped, economic rights. The more a startup equity holder can benefit from channels unrelated to their direct economic rights, the less likely profit-capping will significantly influence their behavior.

A corporate investor primarily interested in early access to technology might push for accelerated R&D while disregarding AI risks, even if their

---

268. *See* Berber Jin et al., *Struggling AI Startups Look for a Bailout from Big Tech*, WALL ST. J. (Aug. 6, 2024), https://on.wsj.com/3WLQFOG; Brian J. Broughman et al., *No Exit*, 100 N.Y.U. L. REV. 1481, 1543–46 (2025).

269. Broughman et al., *supra* note 268, at 1543-46.

270. *Id.*

271. *Id.* at 1543–44.

272. Berber Jin et al., *The Opaque Investment Empire Making OpenAI's Sam Altman Rich*, WALL ST. J. (June 3, 2024), https://on.wsj.com/4didzEA.

273. *Id.*; *see also* Matt Levine, *Musk Chose Who Got Chips*, BLOOMBERG (June 4, 2024), https://bloom.bg/4fERJwx.

274. *See generally* Melvin A. Eisenberg, *Self-Interested Transactions in Corporate Law*, 13 J. CORP. L. 997 (1988).

economic rights are capped. Similarly, a startup equity holder seeking to sell products or services to the startup might encourage business activities that increase their sales, regardless of the implications for AI safety or ethics. Profit-capping may have a limited impact on such investors' behavior; it could only be effective as an interest alignment mechanism where the affected equity holders are primarily motivated by profits and mainly profit through their express economic rights. These assumptions cannot be taken for granted, particularly considering the significant reliance of AI startups on corporate investors.

2. Blindspot No. 2: Profit-Capping and Risk Appetite

Capping equity holders' financial upside could certainly *affect* their expected response to managerial proposals concerning safety and ethics, but not necessarily in the way intended. One of the ideas behind the capped profit structure, as mentioned above, is to limit shareholders' potential upside from value-increasing corporate decisions that would *harm* society's interests.[275] Doing so successfully will necessarily have a net positive effect on society, assuming that the cap is not placed too high to affect the shareholder's expected payouts in any reasonable scenario.[276] Where society's interest and the shareholder's interest diverge, capping the shareholder's financial upside inevitably brings the sum of incentives operating on the shareholder somewhat closer to that of stakeholders who do not enjoy the same financial upside.

However, as further discussed later in this paper, some corporate strategies can serve both purposes at the same time—increasing enterprise value while simultaneously helping to *protect* society from AI risk.[277] Consider, for instance, the development of safety product features that, while reducing the risk of incidents causing harm to society, also increases enterprise value by mitigating the risk of the negative publicity, lawsuits, or undesirable regulatory scrutiny that might follow such safety incidents, for an expected cost that is less than the expected pecuniary benefit.[278] As illustrated below, a profit cap might create a wedge between the societal and pecuniary rationale

---

275. *See supra* Section III.B.

276. A few years ago, criticism of that sort have been made by some with respect to OpenAI's capped profit structure. *See* Gaske, *supra* note 79, at 23–24. OpenAI's recent restructuring, if indeed driven by investor pressure as reported seems to suggest otherwise. *See* discussion *supra* notes 110–11 and accompanying text.

277. *See infra* notes 293–98 and accompanying text.

278. *See infra* notes 293–98 and accompanying text.

for developing the safety feature that would not have been there without the profit cap.

Unlike standard startup shareholders, capped profit holders hold a fixed claim on the startup's economic rights, and similarly to other kinds of fixed claimants, they are expected to be more *risk averse* financially compared to uncapped, residual claimants.[279] This reality would affect the way that capped owners are expected to respond to management's policy on safety and ethics, which, as illustrated below, would be dictated by three key factors: the proposal's expected effect on the startup's enterprise value, the cap amount, and the relevant owners' prosocial tendencies—that is, their willingness to sacrifice their wealth for society's sake. To illustrate how, consider the following hypo.

Alpha is a hypothetical AI startup whose expected value at a future exit, factoring in all possible changes in value (except as described below), is $100 million. Alpha's economic rights are held equally by two shareholders: Romeo, whose holdings are capped and require transfer of any excess to a designated nonprofit, and Juliette, whose holdings are uncapped. Assume that the only way Alpha's equity holders could liquidate their holdings is the exit payouts when Alpha is sold. Assume further that Alpha's managers are committed to AI safety and ethics and have been completely insulated from any personal interests.[280]

Alpha's managers are considering developing a certain safety feature and incorporating it into Alpha's product, which would have a positive societal effect. Assume that each of Romeo and Juliette hold a veto right over all safety-related matters, such that their mutual support is required for such decisions to materialize.[281] The safety feature has a 50% chance of increasing Alpha's exit value to $120 million (for example, by helping mitigate future litigation risk) and a 50% chance of decreasing it to $90 million (e.g., by not mitigating *enough* future litigation risk to make up for the development costs), creating a net positive expected value change of $5 million. As an uncapped shareholder, Juliette's financial interests align directly with Alpha's value. She faces an expected increase of $2.5 million in the value of her 50% stake in Alpha and would support the proposal regardless of her

279. The relationship between shareholders' upside potential and their risk preferences has been extensively explored in various context of corporate law. *See, e.g.*, Easterbrook & Fischel, *supra* note 30, at 403–04; Valesco, *supra* note 30, at 912–16.

280. *See* discussion *supra* Section III.A.

281. *See infra* notes 331–44 for a discussion of the potential use of such veto rights by startup shareholders.

prosocial tendencies. Conversely, as Figure 2 below illustrates, Romeo's incentives vary based on the cap amount.

Figure 2. Expected Value of Romeo's Payouts in Different Cap Scenarios

Profit Cap	Downside Net Payout	Upside Net Payout	Holdings Expected Value (EV)	Δ EV vs. Base
$5,000,000	$5,000,000	$5,000,000	$5,000,000	0
$50,000,000	$45,000,000	$50,000,000	$47,500,000	–$2,500,000
$500,000,000	$45,000,000	$60,000,000	$52,500,000	$2,500,000

With a very low cap (say, $5 million), Romeo becomes financially indifferent to the proposal since his payout remains constant at $5 million across all outcomes. Romeo's gross payouts may range from $45 million to $60 million, but his net payouts are always capped at $5 million. Therefore, his expected value change from moving forward with the proposal is $0. Even minimal prosocial tendencies, however, could lead him to support the proposal.

With a medium cap set near Romeo's expected payout ($50 million), his incentives diverge further from those of Alpha or Juliette. In the upside case, Romeo's share would be worth $60 million but his payout would be capped at $50 million, yielding no additional gain. In the downside case, however, his payout would drop to $45 million, a $5 million loss. Having nothing to gain but something to lose from fluctuations in Alpha's value, Romeo would be *less* supportive of both proposals than Juliette, and will oppose them unless his prosocial tendencies tip the scale the other way.

With a very high cap (say, $500 million), Romeo's incentives mirror Juliette's, as the cap becomes effectively meaningless. Since both the upside and downside scenarios ($45 million or $60 million) fall well below the cap, Romeo is able to enjoy the full upside, and his payoffs match Juliette's. Romeo would support the proposal regardless of the extent of his prosocial tendencies.

What the analysis above illustrates is that profit caps can create unexpected effects. A sky-high cap makes the capped profit structure meaningless and would not affect equity holders' incentives. Lower caps are more likely to affect equity holders' incentives, but may not have the *desired effect* of making them more aligned with society's interests. Where the cap is close enough to the equity holder's expected payouts or below it, they would become less likely than uncapped equity holders to support proposals based on their potential to increase enterprise value.

C. Aligned Choice of Entity

1. "Greedy Shareholder" Problems

a. Safety and Ethics v. Shareholder Primacy

While corporate purpose has been the subject of a century-old debate, it is a widespread perception that Delaware's corporate law follows the shareholder primacy model—meaning, imposing on directors a fiduciary duty of loyalty that requires them to act in the best interests of the corporation's shareholders.[282] Under that perception, directors may generally consider the interests of other constituencies like employees, customers, or society, but those interests should only be considered *instrumentally*, as a means to increase shareholder wealth in the long term.[283]

The duty to maximize shareholder wealth might seem at odds with a solid commitment to AI safety and ethics. In a typical corporate setting, actions to mitigate existential risks, algorithmic biases, or privacy concerns posed by a firm's technology may only be permissible if they are meant to benefit shareholders financially. Prioritizing safety and ethics for the sake of humanity could, theoretically, lead prudent directors into legal trouble, facing accusations from greedy shareholders of neglecting their fiduciary duties.[284]

Concerns about the misalignment between corporate fiduciary duties and commitment to safety and ethics were prominent in both OpenAI and Anthropic's announcements of their new business structures. To emphasize the need for its new structure, Anthropic's announcement first noted how,

---

282. *See* Unocal Corp. v. Mesa Petroleum Co., 493 A.2d 946, 955 (Del. 1985), *holding modified by* Coster v. UIP Cos., Inc., 300 A.3d 656 (Del. 2023); eBay Domestic Holdings, Inc. v. Newmark, 16 A.3d 1, 34 (Del. Ch. 2010); Ann M. Lipton, *What We Talk About When We Talk About Shareholder Primacy*, 69 CASE W. RES. L. REV. 863, 866 (2019).

283. *See* Revlon, Inc. v. MacAndrews & Ford Holdings, Inc., 506 A.2d 173, 182 (Del. 1986); Mills Acquisition Co. v. Macmillan, Inc., 559 A.2d 1261, 1280, 1282 & n.29 (Del. 1989). Where there is no longer a long-term horizon for maximizing shareholder wealth—for instance, where the firm is about to be sold for cash—directors' *Revlon* duties may be triggered, and taking non-shareholder constituents' interests into account becomes generally inappropriate. *See Revlon*, 559 A.2d at 182; *In re* Toys "R" Us, Inc. S'holder Litig., 877 A.2d 975, 999–1000 & n.32 (Del. Ch. 2005); *see also* Robert T. Miller, *Delaware Law Requires Directors to Manage the Corporation for the Benefit of Its Stockholders and the Absurdity of Denying It: Reflections on Professor Bainbridge's* Why We Should Keep Teaching *Dodge v. Ford Motor Co.*, 48 J. CORP. L. 32, 66 (2023); Ann M. Lipton, *Will the Real Shareholder Primacy Please Stand Up?*, 137 HARV. L. REV. 1584, 1595–98 (2024) (reviewing STEPHEN M. BAINBRIDGE, THE PROFIT MOTIVE: DEFENDING SHAREHOLDER VALUE MAXIMIZATION (2023)).

284. *See infra* Section III.C.1.b. for a discussion about the enforceability of shareholder wealth maximization.

under a standard corporate formation, "directors are legally accountable to stockholders for fulfilling their fiduciary duties"[285] and "all the key mechanisms of accountability in corporate law push directors to prioritize the financial interests of stockholders."[286] OpenAI skipped the Business Associations 101 paragraph and went straight to stressing how, under its new structure, its "primary fiduciary obligation is to advance the aims of the OpenAI Charter"[287]—that is, ensuring AGI benefits all humanity,[288] which "always comes first, even at the expense of some or all of [investors and employees'] financial stake."[289]

OpenAI and Anthropic have attempted to resolve this apparent conflict by using alternative business entities that offer prudent directors more freedom to consider non-shareholder constituents' interests. A key aspect of OpenAI's capped profit model was forming OpenAI Global as an LLC.[290] The fiduciary duties owed by LLC members or managers can easily be eliminated by including provisions to that effect in the LLC's governing documents.[291] They may also be expanded or restricted in a similar manner.[292] This adaptability helps insulate prudent managers from fiduciary challenges or create effective mechanisms for shareholders to hold greedy directors accountable. For example, an LLC's operating agreement may provide that LLC managers do not owe fiduciary duties at all or expressly authorize managers to prioritize safety and ethics over members' pecuniary interests. This flexibility has undoubtedly made it easier for OpenAI Global to become "legally bound to pursue the Nonprofit's mission"[293] rather than prioritize the financial interests of OpenAI's investors or employees.

The base layer of Anthropic's model relies on Anthropic's classification as a PBC. Under OpenAI's reformed structure, the for-profit entity is similarly organized as a PBC.[294] PBC directors are required by statute to manage the firm "in a manner that balances the pecuniary interests of the stockholders, the best interests of *those materially affected by the corporation's conduct*, and *the specific public benefit or public benefits*

285. *LTBT Primer*, *supra* note 6.
286. *Id.*
287. *Capped Profit Announcement*, *supra* note 24.
288. *OpenAI Charter*, *supra* note 5.
289. *Capped Profit Announcement*, *supra* note 24.
290. *See* discussion *supra* notes 90–94; *Capped Profit Structure Primer*, *supra* note 13.
291. *See* DEL. CODE ANN. tit. 6, § 18-1101(c) (West 2025).
292. *Id.*
293. *See supra* note 92 and accompanying text.
294. *See* discussion *supra* notes 110–11 and accompanying text.

identified in its certificate of incorporation."[295] Directors' discretion in this respect is protected under a business judgment rule-like standard of review, are deemed to have satisfied their fiduciary duty to balance between these interests where their decision was "both informed and disinterested and not such that no person of ordinary, sound judgment would approve."[296] Anthropic has broadly set its designated public benefit as the responsible development and maintenance of "advanced AI for the long-term benefit of humanity."[297] OpenAI has similarly described its for-profit entity to be "required to advance its stated mission and consider the broader interests of all stakeholders, ensuring the company's mission and commercial success advance together."[298]

b. Can Corporate Directors Take Safety and Ethics Seriously?

While the architects of OpenAI and Anthropic's Aligned Structuring models were apparently dissatisfied with the fit of standard corporations for structuring safety and ethics-conscious startups, corporations can actually be highly protective of safety and ethics-conscious startup directors. Even in a standard corporate setup, startup directors will not necessarily have breached their duties by taking safety and ethics considerations into account, and where they did, the potential plaintiffs would typically lack the evidence, the capacity, or the motivation necessary to have them found liable in court.

First, as mentioned above, directors may generally consider stakeholder interests at least instrumentally, as a means of maximizing long-term shareholder value.[299] In other words, startup directors would not breach their duty of loyalty by taking safety and ethics seriously if they do so believing that unsafe or unethical AI is *bad for business*. It is not hard to imagine why rational startup directors could think so. Safety or ethics-related issues pertaining to a firm's product can aggravate various business risks, from negative publicity and public shaming[300] to litigation with affected

---

295. DEL. CODE ANN. tit. 8, § 365(a) (West 2025) (emphasis added).

296. DEL. CODE ANN. tit. 8, § 365(b) (West 2025).

297. *LTBT Primer*, *supra* note 6.

298. *Our Structure*, *supra* note 97.

299. *See supra* notes 282–93 and accompanying text.

300. OpenAI, for instance, is often fiercely criticized by AI safety advocates for allegedly failing to adequately consider safety in their operations. Sigal Samuel, "*I Lost Trust": Why the OpenAI Team in Charge of Safeguarding Humanity Imploded*, VOX (May 18, 2024), https://www.vox.com/future-perfect/2024/5/17/24158403/openai-resignations-ai-safety-ilya-sutskever-jan-leike-artificial-intelligence [https://perma.cc/KW3Y-ERJS]; Andrew Ross Sorkin et al., *A Safety Check for OpenAI*, N.Y. TIMES (May 20, 2024), https://www.nytimes.com/2024/05/20/business/dealbook/openai-leike-safety-superalignment.html.

stakeholders[301] to increased regulatory enforcement measures.[302] PR crises, hovering lawsuits, or swelling compliance costs could burden a startup's cash-constrained financials, hinder its ability to attract investors, customers, and top talent, or even cripple its business model.[303] While resource-intensive Big Tech firms may be better equipped to weather such storms, it is understandable how startup directors may find their fledgling firms particularly vulnerable and advise caution.

Second, even where shareholders suspect that corporate directors appear to wrongfully prioritize safety and ethics *per se* while disregarding their impact on shareholder wealth, the business judgment rule would usually pose a significant challenge for plaintiffs who may try to bring such an action against them.[304] Before Delaware courts would second-guess the decisions of informed and nonconflicted directors and potentially find them liable, plaintiffs would face the high burden of showing that the directors' actions were the result of *bad faith* (meaning, an intent to harm or an intentional dereliction of their duties)[305] or *wasteful* (meaning, not just unwise but utterly irrational).[306] Indeed, some scholars have suggested that while shareholder wealth maximization may be the applicable standard, the business judgment rule renders it practically unenforceable against directors who prioritize stakeholders' interests instead, unless they have openly admitted to doing

---

301. *See, e.g.*, Douglas B. Laney, *AI Ethics Essentials: Lawsuit Over AI Denial of Healthcare*, FORBES (Nov. 16, 2023), https://www.forbes.com/sites/douglaslaney/2023/11/16/ai-ethics-essentials-lawsuit-over-ai-denial-of-healthcare [https://perma.cc/CFA3-V7W2] (discussing a complaint filed against health insurance carrier UnitedHealthcare for using "nH Predict," an AI model that allegedly terminated coverage prematurely for elderly patients); Katie Robertson, *8 Daily Newspapers Sue OpenAI and Microsoft Over A.I.*, N.Y. TIMES (Apr. 30, 2024), https://www.nytimes.com/2024/04/30/business/media/newspapers-sued-microsoft-openai.html (discussing a lawsuit by newspaper publications accusing OpenAI and Microsoft of illegally using their copyrighted news articles to power their AI-based chatbots).

302. *See, e.g.*, CFPB ET AL., JOINT STATEMENT ON ENFORCEMENT EFFORTS AGAINST DISCRIMINATION AND BIAS IN AUTOMATED SYSTEMS (2023), https://files.consumerfinance.gov/f/documents/cfpb_joint-statement-enforcement-against-discrimination-bias-automated-systems_2023-04.pdf [https://perma.cc/6TPR-3X34].

303. For information about adopting socially responsible strategies as safeguards against similar downside risks, see Stavros Gadinis & Amelia Miazad, *Corporate Law & Social Risk*, 73 VAND. L. REV. 1401, 1448–70 (2020).

304. *See, e.g.*, Edward B. Rock, *Business Purpose and the Objective of the Corporation*, *in* RESEARCH HANDBOOK ON CORPORATE PURPOSE AND PERSONHOOD 27, 41 (Elizabeth Pollman & Robert B. Thompson eds., 2021). For information about the business judgement rule, see DREXLER ET AL., *supra* note 40, at § 15.03.

305. *In re* Walt Disney Co. Derivative Litig., 906 A.2d 27, 67 (Del. 2006); Lyondell Chem. Co. v. Ryan, 970 A.2d 235, 243 (Del. 2009).

306. Glazer v. Zapata Corp., 658 A.2d 176, 183 (Del. Ch. 1993); *In re* Walt Disney Co. Derivative Litig., 731 A.2d 342, 362 (Del. Ch. 1998).

so.[307] To be sure, it would never be lawful—within a corporate structure or otherwise—for managers to *misrepresent* to investors how they balance safety and ethics against shareholder wealth maximization; such conduct could amount, among other things, to a blunt securities fraud.[308] However, where genuine disagreements arise ex post about the weight that should be given to AI safety and ethics, corporate shareholders would find it extremely difficult to contest that balance in court.

Third, private ordering solutions may provide startup directors with even more protection against fiduciary challenges. While the duty of loyalty may not be eliminated through private ordering,[309] Delaware courts have acknowledged that shareholders may covenant not to assert a claim that it was breached.[310] Startup directors could, potentially, require incoming investors to waive all future claims concerning their judgment on safety or ethics issues. Covenants of that sort do have their limits. Their enforceability is subject to a case-by-case review, and they may be invalid if deemed not "narrowly tailored"[311] or unable to "survive close scrutiny for reasonableness."[312] They are also incapable of availing directors from *intentional* breaches of the duty of loyalty[313] and would thus not protect directors who deliberately sacrificed shareholder wealth for safety or ethics' sake. However, covenants not to sue for a breach of fiduciary duties are ordinarily used in startup financing deals—and were proven capable of surviving court scrutiny[314]—in contexts raising comparable questions of

---

307. Christopher M. Bruner, *The Enduring Ambivalence of Corporate Law*, 59 ALA. L. REV. 1385, 1449 (2008); Lynn M. LoPucki, *The End of Shareholder Wealth Maximization*, 56 U.C. DAVIS L. REV. 2017, 2030 (2023). For information about an opposing view, see Miller, *supra* note 283, at 95–104.

308. 17 C.F.R. § 240.10b-5 (2025) provides that:

> It shall be unlawful for any person . . . To make any untrue statement of a material fact or to omit to state a material fact necessary in order to make the statements made, in the light of the circumstances under which they were made, not misleading . . . in connection with the purchase or sale of any security.

309. New Enter. Assocs. 14, L.P. v. Rich, 295 A.3d 520, 581–82 (Del. Ch. 2023) ("[T]he constitutive document of a corporation (the charter and bylaws) (i) can shape fiduciary duties but cannot eliminate them.").

310. *Id.* at 587–89 (discussing Transcript of Record, *In re* Altor Bioscience Corp., C.A. No. 2017-0466-JRS (Del. Ch. May 15, 2019)).

311. "Narrowly tailored" means that the covenant must address a "specific transaction that otherwise would constitute a breach of fiduciary duty." *Id.* at 589.

312. *Id.* at 589–91.

313. *Id.* at 593.

314. *Id.* at 520.

prioritizing competing interests over shareholder wealth maximization (e.g., the fixed, debt-like financial claims of preferred shareholders).[315]

Lastly, even where directors—particularly *entrepreneur-directors*—could technically be found liable for breaching their duties, the relational nature of the entrepreneur-investor relationship makes fiduciary litigation in startups exceedingly rare. The startup ecosystem is a repeat players' market where reputation is a valuable currency.[316] Entrepreneurs or investors with a history of dragging their business partners to court risk their chances of doing business again with other industry players in the future—or at least doing so on favorable terms.[317] VCs, in particular, have been known to be deeply concerned about maintaining their founder-friendly reputation and the adverse effects of confronting entrepreneurs on their future deal flow.[318] As a result, not only are VCs extremely unlikely to sue entrepreneurs,[319] but they are even reluctant to criticize entrepreneurs in public—including entrepreneurs of startups in which they *did not invest*.[320] Anecdotal evidence suggests that VCs are hesitant to do so even in front of the intimate forum present at board meetings.[321]

Admittedly, startup governance disputes can get heated and occasionally do end up in court. OpenAI itself, as mentioned, is currently entangled in litigation over its approach to AI safety and its governance structure.[322]

---

315. Such covenants are ordinarily included as part of drag along provisions and are designed to protect directors from shareholders who might challenge their approval of acquisition proposals, arguing that the directors have wrongly favored the interests of some shareholders or classes over shareholders' common interest. *See NVCA Model Voting Agreement*, NAT'L VENTURE CAP. ASS'N § 3.2(e) (Oct. 2025), https://nvca.org/document/nvca-model-voting-agreement-updated-oct-2025/. For information about drag along agreements, see FELD & MENDELSON, *supra* note 11, at 85–88; HULL ET AL., *supra* note 164, at § 7.25.

316. Pollman, *supra* note 11, at 204.

317. *See* Fried & Ganor, *supra* note 42, at 1001 (discussing entrepreneurs' low likelihood of suing VCs); Broughman & Wansley, *supra* note 42, at 1357 (discussing VCs' low likelihood of suing entrepreneurs).

318. Broughman & Wansley, *supra* note 42, at 1313–17, 1352–58; Pollman, *supra* note 11, at 204–05; Verity Winship, *Unicorn Shareholder Suits*, 100 IND. L.J. 1, 45 (2024).

319. Vladimir Atanasov et al., VCs and the Expropriation of Entrepreneurs 20 (January 2007) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=905923; Broughman & Wansley, *supra* note 42, at 1357.

320. Broughman & Wansley, *supra* note 42, at 1358.

321. *Id.* at 1315.

322. Complaint, Musk v. Altman, No. CGC-24-612746 (Cal. Super. Ct. dismissed June 11, 2024), 2024 WL 899024 (reflecting Musk's first, voluntarily withdrawn suit); *see also* Musk v. OpenAI, Inc., 795 F.Supp.3d 1179 (N.D. Cal. 2025); Bernie Pazanowski, *Musk Accuses Sam Altman of Mishandling OpenAI in New Suit (2)*, BLOOMBERG L. (Aug. 5, 2024), https://news.bloomberglaw.com/litigation/musk-files-new-suit-against-sam-altman-openai-in-federal-court; *see also* discussion *supra* note 107 and accompanying text.

However, generally speaking, litigated governance disputes in startups typically only arise where the different risk appetites, investment horizons, or liquidity needs of the entrepreneurs and various investors are in direct conflict. This usually happens where the startup's *financing* or *exit strategy* are concerned—for instance, when the board approves a new financing round that would wipe out existing investors or a merger that would pay little to low-ranking shareholders.[323] The prospect of startup shareholders suing startup directors over everyday *product-related matters*, whose effect on exit payouts is more speculative and less easily observable, does not naturally fit in Silicon Valley's corporate culture.

2. "Greedy Director" Problems

Some of the comments made in the *Not for Private Gain* letter pointed to the opposite kind of concern: that the choice of entity matters not only for protecting the startup's discretion from harmful interference, but also for enabling desirable interference when necessary to enforce its commitment to AI safety and ethics.[324] They emphasized how "[a] PBC's obligation to its public benefit goal is toothless in practice."[325] While shareholders holding above a minimum holdings threshold may bring legal action in an attempt to enforce the PBC's social mission, such attempts are unlikely and have never been reported in Delaware in fact.[326] Were such attempts to be made, PBC directors' discretion regarding the balancing between shareholder interests, stakeholder interests, and the public purpose is protected under a business-judgment-rule-like standard of review.[327] Even a controlling shareholder in a PBC can—by default—do no more than fire the PBC's directors in case of disagreement,[328] but holds no power to directly interfere with their discretion—as the authority to run a Delaware corporation's business generally rests with the board.[329]

While true, it is important to note that when privately held startups are concerned, the extent of control over the board's decision-making granted to shareholders as a matter of default law has little practical significance.

---

323. *See* Atanasov et al., *supra* note 319, at 20; Pollman, *supra* note 11, at 183.

324. *Not for Private Gain Second Letter*, *supra* note 109, at Part II.

325. *Id.* at Section II.1.

326. *Id.* at Section II.1; DEL. CODE ANN. tit. 8, § 367 (West 2025).

327. *See* discussion *supra* note 296 and accompanying text.

328. *Id.* at Section II.2.

329. *Id.*; *see also* DEL. CODE ANN. tit. 8, § 141(a) ("The business and affairs of every corporation organized under this chapter shall be managed by or under the direction of a board of directors.").

Startups' governance structures rely heavily on private ordering solutions that deviate significantly from default Delaware law.[330] Importantly for the discussion here, it has been a long-standing practice for startups to offer their investors in their charters veto rights over matters that would have otherwise been at the board's sole discretion—for instance, declaring dividends or redeeming share capital.[331] While a certain extent of uncertainty remains about the exact limits of the power to delegate board authorities to shareholders through charter amendments,[332] Delaware law is generally permissive when it comes to private ordering of corporate governance structures and "allows a corporate charter to contain *virtually any provision* related to the corporation's governance" as long as such provisions are not contrary to mandatory Delaware law.[333]

Furthermore, the newly enacted Delaware General Corporation Law § 122(18) expressly authorizes a corporation to enter into contracts with shareholders which, for example, "require the approval or consent of 1 or more persons or bodies before the corporation may take actions specified in the contract," as long as they do not conflict with the provisions of the charter or Delaware law.[334] While case law applying § 122(18) in these circumstances is not yet available, the amendment is widely perceived to allow corporate boards significant freedom in this respect—with some commentators going as far as suggesting that under § 122(18), directors are permitted "to enter into contracts that transfer *all or nearly all of their powers to certain shareholders*."[335] And so, even where a startup was formed as a plain-vanilla Delaware corporation, there appear to be few legal barriers in the way of

---

330. Jill E. Fisch, *Stealth Governance: Shareholder Agreements and Private Ordering*, 99 WASH. U. L. REV. 913, 915–16 (2021).

331. *See, e.g.*, Smith, *supra* note 11, at 346–47. For a standard format of such veto rights, see *NVCA Model COI*, *supra* note 167, at § 3.3.

332. *See* Ann Lipton, *What Is the Value of the Corporate Charter?*, BUS. L. PROF BLOG (May 10, 2024), https://www.businesslawprofessors.com/2024/05/what-is-the-value-of-the-corporate-charter/ [https://perma.cc/VEF5-YBS8] (commenting that some uncertainty remains even after taking into account the provisions of the newly enacted § 122(18)); discussion *infra* notes 334–37 and accompanying text; *see also* Moelis & Co. v. W. Palm Beach Firefighters' Pension Fund, No. 340, 2024, 2026 WL 184868, at *9 (Del. Jan. 20, 2026).

333. Manti Hldgs., LLC v. Authentix Acq. Co., Inc., 261 A.3d 1199, 1217 (Del. 2021) (emphasis added).

334. DEL. CODE ANN. tit. 8, § 122(18) (West 2025).

335. Gabriel Rauterberg & Sarath Sanga, *Altering Rules: The New Frontier for Corporate Governance*, 42 YALE J. ON REG. 291, 322 (2025) (emphasis added); *see also* Jill E. Fisch & Steven Davidoff Solomon, *Contract Rights and Control*, 27 U. PA. J. BUS. L. 101, 117 (2025) ("The exact scope of contractual allegations permitted by section 122(18) has yet to be judicially tested, but the contractual rights which shareholders can now potentially exercise are quite broad.").

affording Alignment Stewards or other shareholders enforcement rights that go far beyond firing the directors or bringing legal action.[336] These powers may, for instance, take the form of protective provisions over board decisions that have significant influence over AI safety or ethics matters—similarly to the method applied in Anthropic's PBC-based structure.[337]

One caveat to the above is that, while on-point case law is yet unavailable, private-ordering-based shareholder oversight powers granted to *prudent* shareholders may be somewhat vulnerable to fiduciary challenges by other, *greedy* shareholders.[338] Even under the new § 122(18), directors cannot afford shareholders oversight power by contract if doing so would itself conflict with their fiduciary duties.[339] And while fiduciary duties cannot excuse non-performance under corporate contracts, directors may also encounter fiduciary challenges if they decide to perform under such shareholder agreements, where their fiduciary duties require that they "efficiently breach" these contracts instead.[340] If corporate directors' fiduciary duties do not allow them to prioritize AI safety and ethics over shareholders' pecuniary interests, affording enforcement power to shareholders who would act against this purpose, or choosing to perform under the agreements providing such powers, might, arguably, constitute a breach of duty, for which greedy shareholders could sue.

That said, as discussed above, greedy shareholders are unlikely to successfully use fiduciary claims to disrupt the creation or performance of these private ordering solutions, just as they are unlikely to do so with respect to other board decisions concerning AI safety and ethics.[341] To challenge a board's decision to enter into such governance agreements ex ante or perform under them ex post, plaintiffs would likely have to overcome at least the

---

336. *Cf.* discussion *supra* notes 326–29 and accompanying text.

337. *See* discussion *supra* note 129 and accompanying text.

338. *See* S. 313, 152d Gen. Assemb. syn. to § 122(18) (Del. 2024). For discussion about the complex (and somewhat uncertain) relationship of new § 122(18) with directors' fiduciary duties, see Fisch & Solomon, *supra* note 335, at 108; Marcel Kahan & Edward Rock, *Proposed DGCL § 122(18), Long-term Investors, and the Hollowing Out of DGCL § 141(a)*, HARV. L. SCH. F. ON CORP. GOVERNANCE (May 21, 2024), https://corpgov.law.harvard.edu/2024/05/21/proposed-dgcl [https://perma.cc/M7X2-XQZ4].

339. *See* sources cited *supra* note 335.

340. *Id.*; *see also* Wagner v. BRP Grp., Inc., 316 A.3d 826, 859–60 (Del. Ch. 2024) (discussing the effect of fiduciary duties on the efficient breach of corporate contracts); Jim An, *Moelis, § 122(18), and Remedies in Contractual Breaches Prompted by Fiduciary Duty*, HARV. L. SCH. F. ON CORP. GOVERNANCE (July 3, 2024), https://corpgov.law.harvard.edu/2024/07/03/moelis [https://perma.cc/39AH-3GYF] (addressing the availability of the remedy of specific performance to the injured shareholder in such circumstances).

341. *See* discussion *supra* Section III.C.1.b.

business judgment rule[342] and reputational constraints,[343] and their ability to do so can be minimized further through contractual solutions to preempt that sort of litigation.[344]

D. The Costs

Even if Aligned Structuring models do not reliably make AI startups more committed to safety or ethics, their use imposes real costs. These costs arise from the fact that Aligned Structuring requires deviations from the startup structuring conventions that are otherwise designed, as discussed above, to optimize the startup's access to capital and talent.[345] Implementing an Aligned Structuring model would necessarily require the startup to compromise on this effort and undermine, at least to some extent, its ability to develop any technology—regardless of whether it is socially desirable or not.

Startups' conventional governance structures are designed to provide entrepreneurs and investors with the clarity and confidence they need to commit their capital—human or financial.[346] Assigning control rights to Alignment Stewards would disrupt this architecture, making investments of human or financial capital riskier and harder to assess than they could have otherwise been. For instance, VCs normally secure a certain extent of control rights over the startup's exit strategy[347]—for example, in order to protect their investments from lowball premature exits[348] or be able to force the exit of underperforming startups and direct their investments to better-performing portfolio assets.[349] If an Alignment Steward is given complete control over the startup's exit strategy to ensure it would not be sold to a greedy acquirer, VCs would likely be more concerned about investing without the extent of exit that their investment thesis is based upon. They could be expected, for instance, to require more favorable economic terms—for example, a lower price per share—for the added risk, thus increasing the startup's financing

---

342. *See* discussion *supra* notes 304–08 and accompanying text.

343. *See* discussion *supra* notes 316–21 and accompanying text.

344. *See* discussion *supra* notes 309–15 and accompanying text.

345. *See* discussion *supra* Part II.

346. *See* discussion *supra* Section II.A.

347. For information about the role of investor control rights in allocating exit control in startups, see Gad Weiss, *The Venture Corporation*, 62 AM. BUS. L.J. 45, 54–55 (2025).

348. *Id.* at 50; *see generally* Matthew Wansley, *Beach Money Exits*, 45 J. CORP. L. 151 (2019).

349. Weiss, *supra* note 347, at 50; In re Trados Inc. S'holder Litig., 73 A.3d 17, 79 (Del. Ch. 2013).

costs.[350] Similarly, entrepreneurs asked to cede control to Alignment Stewards would be rationally more hesitant to invest the required time and effort to develop the startup's product to the best of their ability if they can be replaced at any time based on debatable safety or ethical grounds.[351]

These effects are not limited to the startup's governance structure. Profit-capping mechanisms—while (also) intended to make equity holders less profit-driven[352]—can severely weaken the incentives that typically drive entrepreneurial and investor behavior. Investors, particularly those operating under the power law logic of VC investing, may simply walk away, as a capped investment is less likely to help them "return the fund" as their investment thesis requires.[353] A profit cap might also make a startup lag in the competition over talent, with employees preferring the uncapped equity offered to them elsewhere, or demanding higher base salaries that would leave the startup with less cash for other R&D-related expenses.[354]

A startup's non-optimal choice of entity can impose costs, too. LLCs, while offering great legal flexibility in designing fiduciary duties, may trigger tax and regulatory red flags for VCs.[355] They also require significant deviations from the NVCA-standardized documents investors expect, creating otherwise-avoidable friction and increasing transaction costs when negotiating VC deals.[356] PBCs may be similar to the plain-vanilla corporation and are thus less alien to investors,[357] but offer limited benefits in return—providing very soft constraints on directors' discretion[358] and virtually no enforcement mechanisms for stakeholders that are not available in a standard corporate setup anyway.[359] In this respect, it is important to keep in mind that VCs would ordinarily meet with dozens of potential candidates for each investment they finally make,[360] and the time and attention they can dedicate to each candidate is limited. Startups must avoid standing out *for the wrong reasons*, or they will give other investment candidates a costly competitive

---

350. For information about the relationship between price and non-price terms in VC deals, see Williams, *supra* note 154, at 155–59; Hyde Park Venture Partners Fund III, L.P. v. FairXchange, LLC, No. 2022-0344-JTL, 2024 WL 3579932, at *19 (Del. Ch. July 30, 2024).

351. For information about the effect of the entrepreneurs' control rights over their incentive to make firm-specific human capital investments, see Weiss, *supra* note 347, at 52.

352. *See* discussion *supra* notes 258–60 and accompanying text.

353. *See* discussion *supra* notes 171–72 and accompanying text.

354. *See* discussion *supra* note 173 and accompanying text.

355. *See* discussion *supra* notes 177–87 and accompanying text.

356. *See* discussion *supra* note 188 and accompanying text.

357. *See* discussion *supra* notes 191–208 and accompanying text.

358. *See* discussion *supra* notes 193–208 and accompanying text.

359. *See* discussion *supra* Section III.C.2.

360. Weiss, *supra* note 347, at 63.

edge.[361] For instance, even if the startup will be formed as an LLC but make the necessary "check-the-box" elections to be taxed as a corporation,[362] it would still appear to potential investors as a tax compliance issue at first sight—and thus bear the risk that they move on to the next candidate without ever taking a second look.[363]

Importantly, the costs of using Aligned Structures should not be expected to be uniform. First, they can be expected to be shaped by the robustness of the particular Aligned Structuring model used. A startup that merely gestures at alignment while preserving its true governance and capital structures in the shadows (based on, e.g., informal influence for control[364] and on related-party transactions for financial upside[365]) may enjoy much of the reputational benefit while avoiding many of the structural burdens.

Second, the costs of adopting an Aligned Structuring model would likely also vary with the bargaining power of the parties pushing to do so. For instance, when the model is promoted by a startup's superstar entrepreneurs, investors may gladly overlook even radical deviations from startup financing standards for the privilege to write them a check.[366] It is little surprise that OpenAI and Anthropic, for that matter, could do so as well. With its rare concentration of talent, experience, and star power, [367] early OpenAI was another startup about as much as the 1992 Dream Team was another basketball team.[368] Anthropic's early financing history, exceeding industry benchmark for amounts raised by *ten to twenty times*, similarly indicates their rare and powerful appeal to investors.[369] It is equally unsurprising that as OpenAI's capital needs have grown to levels utterly unprecedented in Silicon

---

361. *Id.*

362. *Id.*

363. *Id.*

364. *See* discussion *supra* Section III.A.1.

365. *See* discussion *supra* notes 272–73 and accompanying text.

366. Serial startups entrepreneurs, and particularly successful ones, were observed to secure favorable investment terms from VCs. *See generally* David H. Hsu, *Experienced Entrepreneurial Founders and Venture Capital Funding*, 36 RES. POL'Y 722 (2007); Rajarishi Nahata, *Success Is Good but Failure Is Not So Bad Either: Serial Entrepreneurs and Venture Capital Contracting*, 58 J. CORP. FIN. 624 (2019).

367. *See* discussion *supra* notes 83–84 and accompanying text.

368. The US 1992 Olympic men's basketball team, commonly referred to as the Dream Team, featured a roster that included, among others, NBA superstars Larry Bird, Michael Jordan, and Magic Johnson. It is considered by many to be the greatest basketball team ever assembled. *See* Joe Vardon, *How USA Basketball Became a Powerhouse, Changed the Olympics and Sparked the WNBA*, ATHLETIC (June 6, 2024), https://nyti.ms/3Wr8wvy.

369. *See* discussion *supra* notes 116–20 and accompanying text.

Valley's history,[370] investors—who, respectively, enjoyed greater bargaining power as OpenAI's capital needs grew[371]—have reportedly become more vocal about their concerns about OpenAI's business structure and pushed for increasingly drastic shifts from its pure nonprofit origins.[372]

For the median startup, however, Aligned Structuring will likely impose significant tradeoffs. And unless those structural choices meaningfully increase a startup's commitment to safety or ethics—which, as discussed above, remains unproven and highly dependent on case-specific facts—these tradeoffs may end up harming not just the startup's founders or investors, but society's interest in access to net-positive AI technology.

IV. THE BOTTOM LINE

For Aligned Structuring to serve as a meaningful tool for reducing the societal risks posed by AI startups, two conditions must be met. First, it must improve decision-making from society's perspective—either by reducing the likelihood of harmful technologies being released or increasing the expected social value of innovation.[373] Second, it must do so without imposing excessive costs on the startup's ability to raise capital, recruit talent, and execute its business plan.[374] This paper's analysis suggests that while Aligned Structuring can satisfy both conditions under specific circumstances, it does not do so reliably or by default. The three main components of Aligned Structuring—Alignment Stewards, profit-capping mechanisms, and alternative entity forms—all present governance tradeoffs that depend heavily on firm-specific facts.

Alignment Stewards are designed to reduce conflict costs by placing decision-making power in the hands of stakeholders with no direct financial interest.[375] But their effectiveness depends on the informal control environment: when startups are highly dependent on their investors, strategic partners, or entrepreneurs, those actors may continue to exert real influence regardless of the extent of formal means of control allocated to them.[376] Alignment Stewards may also increase competence costs in the notional

---

370. Joanna Glasner, *SoftBank and OpenAI Make History with Largest Startup Financing Ever*, CRUNCHBASE NEWS (Apr. 1, 2025), https://news.crunchbase.com/venture/softbank-openai-historic-startup-financing-ai-unicorn/ [https://perma.cc/9M4F-RYCG].

371. *See* discussion *supra* notes 215–18 and accompanying text.

372. *See* discussion *supra* notes 103–10 and accompanying text.

373. *See* discussion *supra* Part III.

374. *See* discussion *supra* Part III.

375. *See* discussion *supra* Section III.A.

376. *See* discussion *supra* Section III.A.1.

agency relationship between society and AI developers if they lack sufficient knowledge or understanding of the startup's technological vision or business model, setting off the benefit derived from reducing conflict costs by assigning control rights to financially disinterested players.[377] However, these drawbacks may be mitigated in specific settings. For instance, where a particular startup is not as reliant on investor capital, entrepreneurs' knowledge assets, or computing power sourced from strategic partners;[378] where the particular Alignment Stewards involved are exceptionally competent,[379] or where the particular technology is relatively mature,[380] requires primarily managers' "policy" rather than "technological" skills to be monitored effectively,[381] or otherwise more easily comprehensible and monitorable by outsiders.

Profit-capping mechanisms face similar limitations. They may reduce the upside incentives that drive reckless or profit-driven behavior, but only when equity holders are genuinely constrained by them. Notably, indirect or alternative profit channels, such as strategic synergies or related-party transactions, can easily dilute the intended effect,[382] and so can high profit caps that exceed financially interested stakeholders' profit expectations.[383] Yet again, the effectiveness of the aligned structure would vary based on firm-specific facts. Where investors are primarily financial rather than strategic[384] and have no other business ties to the startup,[385] and where the profit caps are not too high for financially interested stakeholders to care about,[386] profit caps may become more effective. However, even when capable of effectively affecting financially interested stakeholders' incentives, profit caps may backfire by inducing risk aversion of the unwelcomed kind and curbing the economic incentive to invest in socially beneficial product features.[387]

Alternative entity forms, such as PBCs and LLCs, are assumed by some to improve interest alignment by changing directors' fiduciary duties to allow greater emphasis on societal interests[388] or by expanding stakeholder

377. *See* discussion *supra* Section III.A.1.
378. *See* discussion *supra* notes 215–28 and accompanying text.
379. *See* discussion *supra* notes 243–44 and accompanying text.
380. *See* discussion *supra* notes 245–48 and accompanying text.
381. *See* discussion *supra* notes 256–57 and accompanying text.
382. *See* discussion *supra* Section III.B.1.
383. *See* discussion *supra* note 276 and accompanying text.
384. *See* discussion *supra* notes 262–71 and accompanying text.
385. *See* discussion *supra* notes 272–74 and accompanying text.
386. *See* discussion *supra* note 276 and accompanying text.
387. *See* discussion *supra* Section III.B.1.
388. *See* discussion *supra* notes 285–97 and accompanying text.

enforcement of such duties.[389] But the standard Delaware corporation already provides robust protection to corporate directors from interference by "greedy shareholders." They can often justify safety and ethics as value-maximizing,[390] the business judgment rule would often preempt substantive second-guessing by courts of directors discretion,[391] private ordering tools can further limit exposure,[392] and reputational norms disincentivize shareholder litigation in startups.[393] In "greedy director" cases, shareholder ability to enforce AI safety and ethics-related commitments can be effectively addressed through private ordering solutions without abandoning the standard corporate form.[394] In both scenarios, alternative entities offer, at best, marginal additional value.

Aligned Structuring models also come with an inherent set of real costs.[395] Deviating from standard governance and capital structures can chill investor interest and drive up financing costs,[396] deter entrepreneurial effort,[397] and place the startup in an inferior position in the competition over top talent.[398] Using weaker forms of Aligned Structuring may help avoid or minimize these added costs, but only by watering down the model's intended effect.[399] Aligned Structuring models can also be more effective when adopted by celebrity startups who enjoy unusually significant bargaining power when negotiating with potential investors and employees.[400]

The bottom line is this: Aligned Structuring is not a defective concept, but it is a very fragile one. Its benefits are highly contingent, and its expected costs are potentially significant—for the startup and its owners and financially interested stakeholders, but also for society that will suffer the costs of delayed AI innovation of any kind, regardless of its risk profile. There is, therefore, no reliable reason to assume that an "Aligned" startup is more aligned in fact with society's interest, all things considered—and certainly no basis to treat such structuring as a governance panacea or public policy benchmark.

---

389. *See* discussion *supra* notes 324–29 and accompanying text.
390. *See* discussion *supra* notes 299–303 and accompanying text.
391. *See* discussion *supra* notes 304–08 and accompanying text.
392. *See* discussion *supra* notes 309–15 and accompanying text.
393. *See* discussion *supra* notes 316–23 and accompanying text.
394. *See* discussion *supra* Section III.C.2.
395. *See* discussion *supra* Section III.D.
396. *See, e.g.*, discussion *supra* notes 347–50, 352–53 and accompanying text.
397. *See, e.g.*, discussion *supra* note 351 and accompanying text.
398. *See, e.g.*, discussion *supra* note 354 and accompanying text.
399. *See* discussion *supra* notes 364–65 and accompanying text.
400. *See* discussion *supra* notes 366–72 and accompanying text.

While this paper does not take a position on the merits of OpenAI's recent restructuring,[401] its bottom line speaks directly to the kind of concerns raised in the *Not for Private Gain* letter.[402] The letter drafters urged the Attorneys General of California and Delaware to intervene based on the belief that OpenAI's distinctive structure was essential to safeguarding the public interest.[403] But as the analysis here shows, focusing on a startup's formal business structure—its governance structures, capital structures, and choice of entity—offers an incomplete and potentially misleading picture. Policymakers and regulators seeking to protect the public from AI risk must look beyond surface-level features of a startup's on-paper business structure and ask harder, fact-sensitive questions about how control and cash flows are allocated in fact.

V. CONCLUSION

This paper set out to test whether the three signature elements of Aligned Structuring—(1) handing formal control to Alignment Stewards, (2) capping investors' upside, and (3) adopting alternative legal entities—can reliably steer AI startups toward protecting the public interest without stripping them of the basic ingredients that fuel innovation in venture-backed startups. It concludes that because informal power levers (cash, computing power, talent) often swamp the startup's formal guardrails, profit caps can be side-stepped or induce excessive risk-aversion, and alternative business entities add only marginal legal protection of society's interest, these models' benefits are only expected to surface under narrow, fact-specific conditions. At the same time, as Aligned Structuring deviates from traditional venture-financing norms, they make "aligned" startup equity less attractive. The expected result is a thinner pipeline of capital, talent, and entrepreneurial energy in Aligned startups—and, with it, a slower pace of potentially welfare-enhancing AI advances, unless the startup is a high-profile "celebrity" startup that can readily secure resources on favorable terms, or where the impact of the Aligned Structuring model applied is low.

Aligned Structuring is therefore a fragile concept, even if not faulty by design. It can be effective in protecting society's interest in having access to advanced AI technologies that are beneficial and safe when everything lines up, but it may just as easily dampen innovation without improving safety to a significant degree or do neither and be effectively meaningless.

---

401. *See* discussion *supra* notes 103–10 and accompanying text.
402. *Not For Private Gain First Letter*, *supra* note 1.
403. *Id.*

These findings speak directly to the worries voiced in the *Not for Private Gain* letter, which urged state Attorneys General to block OpenAI's restructuring on the theory that its bespoke governance safeguards were indispensable. The analysis here explains why structural guardrails alone cannot bear that regulatory weight. Choice of entity, profit capping, and the allocations of voting power and board seats matter, but taken at face value, they are poor proxies for where real power and incentives lie in a venture ecosystem that is both privately ordered and relentlessly shaped by informal means of control. Regulators who focus only on business structure risk mistaking form for substance.