CHAPTER FOUR

AMORAL DRIFT IN AI CORPORATE GOVERNANCE

ChatGPT's debut in November of 2022 set off a race in Silicon Valley to develop and monetize artificial intelligence (AI).[1] Within a few months, Microsoft invested $10 billion in OpenAI, the company behind ChatGPT.[2] Anthropic, a competitor of OpenAI, raised similarly impressive amounts of money from companies and investors hoping to participate in the AI revolution.[3]

Well before ChatGPT emerged, commentators warned of the risks advanced AI might pose.[4] Observers who predict existential threats to humanity from superintelligent AI point to the difficulty of precisely controlling it.[5] They reason that superintelligent AI might pursue a human-directed goal without balancing its goal against general human values.[6] For example, with access to enough tools, a superintelligent AI instructed to maximize paperclip production might end up "converting . . . large chunks of the observable universe into paperclips."[7] Alternatively, a superintelligent AI may develop its own unexpected goals — goals that do not necessarily account for human wellbeing.[8] The proposed solution to these types of existential AI risks is "AI alignment": the challenging task of ensuring that the values of an AI align with human values.[9] Critics believe AI startups are moving much faster than AI alignment research can keep up, at great risk to humanity.[10]

---

[1] *See* Karen Weise et al., *Inside the A.I. Arms Race that Changed Silicon Valley Forever*, N.Y. TIMES (Sept. 25, 2024), https://www.nytimes.com/2023/12/05/technology/ai-chatgpt-google-meta.html [https://perma.cc/RP25-BNHT].

[2] Dina Bass, *Microsoft Invests $10 Billion in ChatGPT Maker OpenAI*, BLOOMBERG (Jan. 23, 2023, 5:03 PM), https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai [https://perma.cc/95T2-C78X].

[3] 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://www.nytimes.com/2024/02/20/technology/anthropic-funding-ai.html [https://perma.cc/VV9S-Y3NE].

[4] *See, e.g.*, Scott Alexander, *Superintelligence FAQ*, LESSWRONG (Sept. 20, 2016), https://www.lesswrong.com/posts/LTtNXM9shNM9AC2mp/superintelligence-faq [https://perma.cc/JJ5L-YUSR]; NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES (2014).

[5] *See* Alexander, *supra* note 4.

[6] Scott Alexander, *Why I Am Not (As Much of) a Doomer (As Some People)*, SUBSTACK: ASTRAL CODEX TEN (Mar. 14, 2023), https://www.astralcodexten.com/p/why-i-am-not-as-much-of-a-doomer [https://perma.cc/CNR9-WU48].

[7] BOSTROM, *supra* note 4, at 123.

[8] *See generally* Paul Christiano, *What Failure Looks Like*, LESSWRONG (Mar. 17, 2019, 4:18 PM), https://www.lesswrong.com/posts/HBxe6wdjxK239zajf/what-failure-looks-like [https://perma.cc/AR53-QT4H] (describing how "influence-seeking" AIs might arise and take control).

[9] Sean Welsh, *"Superintelligence," Ten Years On*, QUILLETTE (July 2, 2024), https://quillette.com/2024/07/02/superintelligence-10-years-on-nick-bostrom-ai-safety-agi [https://perma.cc/39NY-K9W7].

[10] *E.g.*, Eliezer Yudkowsky, Opinion, *Pausing AI Developments Isn't Enough. We Need to Shut It All Down*, TIME (Mar. 29, 2023, 6:01 PM), https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough [https://perma.cc/62YK-C5CY].

Even if these existential risks sound far-fetched, AI certainly does present a challenge to existing legal and social frameworks. Companies have already demonstrated that AI can learn from and reflect human racial and gender biases.[11] Both the training inputs and the creative outputs of AI raise complicated questions of intellectual property law.[12] The current spotlight on AI also brings into focus the question of how to protect privacy in the era of Big Data,[13] especially as AI promises to massively boost data collection.[14] More gravely, malicious actors might use AI for terrorism, disinformation, and oppression.[15] AI startups need to confront these legal, ethical, and security issues implicated by AI as they advance the technology, including whether and how to implement guardrails to prevent the misuse of their products.

The risks posed by AI development have revived the question of how to deal with the negative externalities of corporations. Doubtful of the traditional profit motive, AI company founders have adopted some of the most ambitious versions of "prosocial" corporate governance mechanisms detailed in corporate governance literature.[16] To counterbalance the pressure to maximize profit, OpenAI and Anthropic have granted their boards outsized discretion in a manner consistent with a stakeholderist corporate governance mandate. The stakeholder-focused view pushes for the board to consider in its decisionmaking process constituencies other than shareholders, such as consumers, employees, the surrounding community, or even the environment.[17]

---

[11] *See supra* ch. I, pp. 1562, 1573–74.

[12] *See supra* ch. II, p. 1591; Note, *Recovering Personality in Copyright's Originality Inquiry*, 138 HARV. L. REV. 1123, 1127–28 (2025) (discussing the U.S. Copyright Office's refusal to grant a copyright for an AI-created work).

[13] Isabel Gottlieb & Cassandre Coyer, *AI's Data Appetite Is Huge. That's a Problem for Privacy Laws*, BLOOMBERG L. (July 24, 2024, 5:03 AM), https://news.bloomberglaw.com/artificial-intelligence/ais-data-appetite-is-huge-thats-a-problem-for-privacy-laws [https://perma.cc/5M3N-4ACE].

[14] *See, e.g.*, Kashmir Hill, *The Secretive Company that Might End Privacy as We Know It*, N.Y. TIMES (Nov. 2, 2021), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html [https://perma.cc/CC8N-WLNB].

[15] *See* Scott Alexander, *Updated Look at Long-Term AI Risks*, SUBSTACK: ASTRAL CODEX TEN (July 30, 2021), https://www.astralcodexten.com/p/updated-look-at-long-term-ai-risks [https://perma.cc/Y952-873P]; *see also, e.g.*, OPENAI, INFLUENCE AND CYBER OPERATIONS: AN UPDATE 14–19 (2024), https://cdn.openai.com/threat-intelligence-reports/influence-and-cyber-operations-an-update_October-2024.pdf [https://perma.cc/TG5K-4G86].

[16] *See generally* Gad Weiss, Aligned Structuring of AI Startups (Sept. 2, 2024) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4943590 [https://perma.cc/WE9W-P8M2]. Consistent with the broader literature, this Chapter uses the term "prosocial" to describe market actors that seek to pursue social welfare objectives in spite of reduced financial gain. *See, e.g.*, Dorothy S. Lund, Essay, *Corporate Finance for Social Good*, 121 COLUM. L. REV. 1617, 1622 n.17 (2021).

[17] This definition understates the wide range of stakeholderist theories as well as the robust, longstanding debates on corporate purpose occurring in the academic literature and practitioner landscape. *Compare, e.g.*, Lucian A. Bebchuk & Roberto Tallarita, *The Illusory Promise of Stakeholder Governance*, 106 CORNELL L. REV 91, 94, 114, 117 (2020), *with* Martin Lipton, *It's Time to*

These innovations are not perfect. The drama that played out at OpenAI, where powerful investor-supplier Microsoft and an irreplaceable labor force successfully reinstated Sam Altman after he was fired by the nonprofit board,[18] suggests that OpenAI may already have drifted substantially from its initial commitments despite its novel structure. The ability of Anthropic to hold on to its mission of safe AI development remains to be seen, but OpenAI's form of "drift" has long been contemplated in corporate governance and organizational business literature.[19] This mission drift — termed "amoral drift" by Professors Oliver Hart and Luigi Zingales — predicts the slow death of prosocial corporate missions as a result of market pressures and lax legal guardrails.[20]

By weakening shareholders, have the AI companies been solving for the wrong variable? Corporate governance innovations like stakeholderist mandates respond to shareholder-led threats like lawsuits, activist campaigns, and director elections.[21] Such orientation is likely a natural result of orthodox corporate governance theory focusing on shareholders.[22] These corporate governance innovations also tend to focus on public or soon-to-be-public companies.[23] But here, surrounding constituencies in the American tech environment, not traditional shareholders, may be the biggest threat to private AI companies' missions. Given the novelty of this dynamic, insufficient attention has been paid to the maintenance of prosocial aims in private, closely held companies. More novel still, the supercharged development of the AI industry invites the

---

Adopt the New Paradigm, HARV. L. SCH. F. ON CORP. GOVERNANCE (Feb. 11, 2019), https://corpgov.law.harvard.edu/2019/02/11/its-time-to-adopt-the-new-paradigm [https://perma.cc/Q34F-ZZZP], *and* William Savitt & Aneil Kovvali, Essay, *On the Promise of Stakeholder Governance: A Response to Bebchuk and Tallarita*, 106 CORNELL L. REV. 1881, 1883 (2021).

[18] *See* Clare Duffy & Diksha Madhok, *OpenAI's Wild Week. How the Sam Altman Story Unfolded*, CNN (Nov. 22, 2023, 3:32 PM) https://www.cnn.com/2023/11/22/tech/openai-sam-altman-chaos-explained-intl-hnk/index.html [https://perma.cc/VS7J-PNQH].

[19] *See, e.g.*, Einer Elhauge, *Sacrificing Corporate Profits in the Public Interest*, 80 N.Y.U. L. REV. 733, 735–45 (2005); Burton A. Weisbrod, *The Pitfalls of Profits*, 2 STAN. SOC. INNOVATION REV., Winter 2004, at 40, 44; Alnoor Ebrahim et al., *The Governance of Social Enterprises: Mission Drift and Accountability Challenges in Hybrid Organizations*, 34 RSCH. ORGANIZATIONAL BEHAV. 81, 82 (2014).

[20] *See* Oliver Hart & Luigi Zingales, *Companies Should Maximize Shareholder Welfare Not Market Value*, 2 J.L. FIN. & ACCT. 247, 255–56 (2017). The concept of amoral drift was first applied to AI companies by Professor Roberto Tallarita. *See* Roberto Tallarita, *AI Is Testing 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/S7SC-QSJV].

[21] *See* Bebchuk & Tallarita, *supra* note 17, at 112, 165.

[22] This Chapter observes that as a general, descriptive matter, corporate boards appear primarily to pursue value for their shareholders. *See* Hart & Zingales, *supra* note 20, at 257; Dorothy S. Lund & Elizabeth Pollman, Essay, *The Corporate Governance Machine*, 121 COLUM. L. REV. 2563, 2565 (2021). *But see Business Roundtable Redefines the Purpose of a Corporation to Promote "an Economy that Serves All Americans,"* BUS. ROUNDTABLE (Aug. 19, 2019) [hereinafter *Business Roundtable*], https://www.businessroundtable.org/business-roundtable-redefines-the-purpose-of-a-corporation-to-promote-an-economy-that-serves-all-americans [https://perma.cc/N5XP-7HKR].

[23] *Cf.* Elizabeth Pollman, *Private Company Lies*, 109 GEO. L.J. 353, 358–59 (2020).

question of whether prevailing understandings of private companies and of socially oriented companies can even be applied to AI at all.

This Chapter seeks to address why corporate governance tools have failed, and will continue to fail, in preventing amoral drift in companies like OpenAI and Anthropic. Despite attempts to constrain shareholders' ability to orient the firms toward profit, OpenAI and Anthropic will likely still experience amoral drift toward profit maximization due to the unconstrained power of stakeholders. Specifically, "superstakeholders" — stakeholders given significant stakes in the startup's future profits — drive these AI companies to maximize profit. In other words, OpenAI and Anthropic shifted influence away from profit-focused shareholders but left their missions vulnerable to pressure from profit-focused superstakeholders. The superstakeholder problem particularly troubles AI startups, which rely on scarce talent and Big Tech resources to an uncommon degree.[24] For AI companies to preserve their prosocial mission and prevent amoral drift, they must focus their energies on all equity-compensated actors rather than just traditional stockholders.

Section A explores "amoral drift" as a launching point, explaining how AI companies may frustrate the theory's original assumptions. Section B provides an overview of Anthropic's corporate governance model and OpenAI's corporate governance model. Section C describes the various shareholders and stakeholders playing a role in AI corporate governance. Section D assesses specific facets of OpenAI and Anthropic's governance structure to suggest that such structures may not be well equipped for ensuring AI safety. The last section concludes.

A. Amoral Drift

U.S.-focused corporate law scholars have long debated the purpose of corporations and their role in reining in the externalities they impose on the rest of society.[25] While this debate has been longstanding, efforts to make corporations directly responsible for their own externalities through modern corporate governance emerged in the late twentieth century.[26] In the wake of highly publicized corporate misconduct, economic crisis, and legislative dysfunction,[27] corporate governance grew to focus on the "balance of power among shareholders, boards of directors, and managers"[28] rather than on the permissions granted through

---

[24] *See* John Thornhill, *How Big Tech Is Winning the AI Talent War*, FIN. TIMES (Mar. 22, 2024), https://www.ft.com/content/2892bac2-d848-49ea-b983-bc649a8c0529 [https://perma.cc/KGP6-3RVR].

[25] *See generally, e.g.*, E. Merrick Dodd, Jr., *For Whom Are Corporate Managers Trustees?*, 45 HARV. L. REV. 1145 (1932); A.A. Berle, Jr., *For Whom Corporate Managers Are Trustees: A Note*, 45 HARV. L. REV. 1365 (1932).

[26] *See* Mariana Pargendler, *The Corporate Governance Obsession*, 42 J. CORP. L. 359, 366–67 (2016).

[27] *See id.* at 367, 373–74.

[28] *Id.* at 362.

the corporate charter.[29] Following this shift in focus, corporate governance became an attractive avenue for pursuing social change and economic growth simultaneously,[30] and the concept of stakeholderism — calling for corporations to consider not just their shareholders, but also other groups affected by the corporation's actions — arose.[31] The stakeholderism movement reached a high point in 2019 when the Business Roundtable endorsed stakeholderism in its Statement on the Purpose of the Corporation, demonstrating the receptiveness of business leaders to stakeholderism at the time.[32] Emblematic of this strand of thought (though not singular in their view) are Professors Oliver Hart and Luigi Zingales, who have argued that corporations need not and ought not focus exclusively on maximizing profit for shareholders.[33] "Amoral drift," the process through which market-driven preoccupation with stock price forces corporate managers to abandon social concerns,[34] offers a theory to justify stakeholder-centric corporate governance reforms. This section provides an overview of "amoral drift" theory, which this Chapter uses as a foundation for inquiry into private AI companies.

In their influential[35] paper *Companies Should Maximize Shareholder Welfare Not Market Value*, Hart and Zingales reject Milton Friedman's well-known argument that corporations should pursue only profit.[36] For one thing, the government cannot perfectly use regulation to force corporations to internalize all externalities.[37] Moreover, the solution to a negative externality is not always perfectly separable from the externality-causing activity: For example, not burning coal in the

---

[29] *See id.* at 362–63, 374.

[30] *See id.* at 366–67.

[31] *See* Bebchuk & Tallarita, *supra* note 17, at 104–05.

[32] *See Business Roundtable*, *supra* note 22. Since this press release, there has been significant backlash to stakeholderism, most notably through the "anti-ESG" movement, which has led many companies to walk back their commitments to ESG (environmental, social, governance). *Compare, e.g.*, *Larry Fink's 2021 Letter to CEOs*, BLACKROCK, https://www.blackrock.com/us/individual/2021-larry-fink-ceo-letter [https://perma.cc/SA97-2SKN], *and Larry Fink's 2022 Letter to CEOs: The Power of Capitalism*, BLACKROCK, https://www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter [https://perma.cc/QFU4-BRHE], *with* Catherine Clifford, *Larry Fink: BlackRock Is Not the "Environmental Police,"* CNBC (Mar. 15, 2023, 12:21 PM), https://www.cnbc.com/2023/03/15/larry-fink-blackrock-is-not-the-environmental-police-.html [https://perma.cc/TGQ2-GAZM].

[33] *See* Hart & Zingales, *supra* note 20, at 248.

[34] *See id.* at 254.

[35] *See, e.g.*, Lund, *supra* note 16, at 1627–28, 1628 n.39 (citing Hart & Zingales); Kobi Kastiel & Yaron Nili, *The Corporate Governance Gap*, 131 YALE L.J. 782, 859 n.308 (2022) (same); Marcel Kahan & Edward Rock, *Corporate Governance Welfarism*, 15 J. LEGAL ANALYSIS 108, 112 (2023) (discussing Hart & Zingales); Adi Libson, *Taking Shareholders' Social Preferences Seriously: Confronting a New Agency Problem*, 9 U.C. IRVINE L. REV. 699, 700–07 (2019) (same).

[36] *See* Hart & Zingales, *supra* note 20, at 247–48; Milton Friedman, *A Friedman Doctrine — The Social Responsibility of Business Is to Increase Its Profits*, N.Y. TIMES (Sept. 13, 1970), https://www.nytimes.com/1970/09/13/archives/a-friedman-doctrine-the-social-responsibility-of-business-is-to.html [https://perma.cc/VQ9Y-VZYM].

[37] *See* Hart & Zingales, *supra* note 20, at 249.

first place is a more efficient way to reduce pollution than scrubbing the pollution from the atmosphere after the fact.[38] In light of these realities, Hart and Zingales believe companies cannot simply focus on profit; they should consider "shareholder welfare" — that is, the preferences expressed by shareholders as whole individuals rather than as purely profit-maximizing owners of the company.[39] The authors introduce "amoral drift": the tendency of public companies to shed prosocial commitments over time because the risk of corporate takeover and incorrect perceptions about fiduciary duties lead boards to choose profit-maximizing corporate actions.[40] The introduction of this concept builds on a lengthy strand of social enterprise scholarship expressing concern about the increasing number of nonprofits pursuing commercial activities,[41] as well as previous corporate law scholarship rejecting a narrow profit-maximization focus but highlighting its inevitability due to market pressure.[42]

Hart and Zingales posit that shareholders tend to be prosocial only to the extent that they "feel[] responsible for the [dirty] action in question."[43] As a result, if a bidder approaches a corporation with a tender offer, claiming to boost profits by making the company "dirty," a shareholder will weigh the social damage caused by the "dirty" bidder, the price of the tender offer, and the extent to which his vote will determine the outcome of the tender offer.[44] Because in public, widely held corporations a single share's voting power is negligible, a single shareholder may not feel responsible for the outcome, leading prosocial shareholders to tender to a "dirty" bidder even if they would prefer the company to be "clean."[45] Boards that wish to maintain control will adopt profit-boosting, "dirty" corporate strategies so as not to be bested by "dirty" bidders.[46] Furthermore, boards "think . . . that they have a fiduciary duty to maximize shareholder value," so the result holds even in the absence of credible bidders.[47]

Hart and Zingales provide potential approaches for stemming the tide of amoral drift, assuming that the founder is looking for a way to prevent amoral drift and keep her company "clean."[48] Her options include implementing protective measures sanctioned by Delaware law, such as "clean" charter provisions, dual-class shares with unequal voting

---

38 *See id.*
39 *See id.* at 248.
40 *Id.* at 255–58.
41 *See, e.g.*, *supra* note 19 and accompanying text.
42 *See, e.g.*, *supra* note 19 and accompanying text.
43 Hart & Zingales, *supra* note 20, at 253 (emphasis omitted).
44 *See id.* at 255–56.
45 *See id.*
46 *See id.* at 256.
47 *Id.* at 257 (emphasis omitted).
48 *See id.* at 259–60.

rights, and entrenched charitable foundation–style boards.[49] All of these options involve wresting power from the shareholders and redistributing that power to the board or a controlling shareholder.[50]

Hart and Zingales's conception of amoral drift represents a prevailing strand of thought in socially minded corporate governance innovation: that rational, profit-focused shareholders reacting to market pressures are the primary threat to a corporation's social mission.[51] Stakeholderists often lament boards' invocation of shareholder pressure as a way of shirking social commitments.[52] This perspective, in its traditional presentation, tends to position orthodox shareholders against stakeholders, who are stereotypically the victims of corporate externalities and the beneficiaries of corporate social goals. Scholars have explored board-protective measures[53] to ensure consideration of all stakeholders in board decisionmaking,[54] and OpenAI and Anthropic are some of the latest firms to experiment in this area of corporate governance.

B. Anthropic and OpenAI

In light of the well-recognized risks, AI startups — Anthropic and OpenAI — have arranged novel corporate governance structures. As their thinking went, the profit motive is inadequate for policing the risks AI products might pose.[55] Consequently, these startups eschewed the traditional setup for American corporations.[56] In the typical corporation, a board of directors elected by shareholders oversees the company

---

[49] *See id.*

[50] *See id.* Hart and Zingales take an institutionalist approach to corporate social missions, emphasizing that bolstering managerial power can prevent amoral drift. But many scholars have argued for a more robust reexamination of corporate governance institutions that prop up corporate apathy to social welfare. *See generally* Lund & Pollman, *supra* note 22 (highlighting how a broader "system . . . composed of law, institutions, and culture" reinforces shareholder primacy, *id.* at 2565); Emilie Aguirre, *Beyond Profit*, 54 U.C. DAVIS L. REV 2077 (2021) (calling for the enactment of a voluntary commitment mechanism in corporate law); LYNN STOUT, THE SHAREHOLDER VALUE MYTH 3–4 (2012).

[51] *See, e.g.*, Elhauge, *supra* note 19, at 742; Lucian Bebchuk & Oliver Hart, *Takeover Bids vs. Proxy Fights in Contests for Corporate Control* 10 (Nat'l Bureau of Econ. Rsch., Working Paper No. 8633, 2001); *see also* STOUT, *supra* note 50, at 9–10 ("[Shareholder primacy theory] reduces investors to their lowest possible common human (or perhaps subhuman) denominator: impatient, opportunistic, self-destructive, and psychopathically indifferent to others' welfare.").

[52] *See, e.g.*, Lipton, *supra* note 17.

[53] *See generally* Dhruv Aggarwal et al., *The Rise of Dual-Class Stock IPOs* 2–3 (Nat'l Bureau of Econ. Rsch., Working Paper No. 28609, 2021); Richard D. MacMinn & Douglas O. Cook, *An Anatomy of the Poison Pill*, 12 MANAGERIAL & DECISION ECON. 481 (1991) (describing the operation of poison pills as takeover defenses); Yakov Amihud et al., *Settling the Staggered Board Debate*, 166 U. PA. L. REV. 1475 (2018) (describing staggered boards and their effects on firm value).

[54] *See* Caley Petrucci & Guhan Subramanian, *Pills in a World of Activism and ESG*, 1 U. CHI. BUS. L. REV. 417, 424 (2022).

[55] *See Our Structure*, OPENAI, https://openai.com/our-structure [https://perma.cc/R3LR-FM9E].

[56] *See id.*; *The Long-Term Benefit Trust*, ANTHROPIC (Sept. 19, 2023), https://www.anthropic.com/news/the-long-term-benefit-trust [https://perma.cc/3CQK-DD8P].

with the goal of "maximiz[ing] value" for the shareholders.[57] Anthropic and OpenAI instead devised ways of insulating their boards from shareholder pressure, on the theory that an insulated board would make safer, more socially responsible decisions.[58] This section examines those methods in practice and as contrasted with academic theory.

1. Anthropic. — Anthropic combines a rarely used form, the public benefit corporation (PBC), with a novel structure that it calls the "Long-Term Benefit Trust."[59] The board of a typical corporation owes fiduciary duties that run to shareholders, who attempt to enforce these duties through shareholder votes and lawsuits.[60] In a Delaware PBC, the board must "balance[] the pecuniary interests of the stockholders, the best interests of those materially affected by the corporation's conduct, and the specific public benefit . . . identified in [the] certificate of incorporation."[61] Anthropic has identified its "public benefit purpose" as "the responsible development and maintenance of advanced AI for the long-term benefit of humanity."[62]

Seeking to further shield the board from potential shareholder pressure, Anthropic built a mechanism into its charter that empowers safety-focused trustees. After May 24, 2027, or eight months after obtaining a total of $6 billion in investments — whichever comes first — "Class T" shareholders alone will elect three of Anthropic's five board directors.[63] Preferred stockholders (usually venture capitalists and other investors[64]) and common stockholders (typically founders and employees[65]) will elect one director each.[66] The trust holds all of the Class T shares, meaning it will eventually control a majority of the board.[67] It appears Anthropic cleared the $6 billion hurdle by the first half of 2024,[68] so the trust will gain control in 2025 at the latest.

---

57 Ann M. Lipton, *Will the Real Shareholder Primacy Please Stand Up?*, 137 HARV. L. REV. 1584, 1592 & n.55 (2024) (book review) (quoting Frederick Hsu Living Tr. v. ODN Holding Corp., No. 12108, 2017 WL 1437308, at *20 (Del. Ch. Apr. 14, 2017)); *id.* at 1588. *See generally* Lund & Pollman, *supra* note 22.

58 *See Our Structure*, *supra* note 55; *The Long-Term Benefit Trust*, *supra* note 56.

59 *The Long-Term Benefit Trust*, *supra* note 56.

60 *See* Lipton, *supra* note 57, at 1588.

61 DEL. CODE ANN. tit. 8, § 365 (2024).

62 *The Long-Term Benefit Trust*, *supra* note 56.

63 Anthropic, PBC, Amended and Restated Certificate of Incorporation 21–22 (May 20, 2024).

64 *See* Elizabeth Pollman, *Startup Governance*, 168 U. PA. L. REV. 155, 160, 173–74, 174 n.96 (2019); *see also* Dylan Matthews, *The $1 Billion Gamble to Ensure AI Doesn't Destroy Humanity*, VOX (Sept. 25, 2023, 10:30 AM), https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2 [https://perma.cc/2YFY-5WEA].

65 *See* Pollman, *supra* note 64, at 160.

66 Anthropic, *supra* note 63, at 21.

67 John Morley et al., *Anthropic Long-Term Benefit Trust*, HARV. L. SCH. F. ON CORP. GOVERNANCE (Oct. 28, 2023), https://corpgov.law.harvard.edu/2023/10/28/anthropic-long-term-benefit-trust [https://perma.cc/5EJF-BDL5].

68 *See* Billy Perrigo, *How Anthropic Designed Itself to Avoid OpenAI's Mistakes*, TIME (May 30, 2024, 1:46 PM), https://time.com/6983420/anthropic-structure-openai-incentives [https://perma.cc/B79N-8DEJ].

Anthropic and its lawyers describe the Long-Term Benefit Trust as a Delaware common law purpose trust.[69] A typical common law trust is organized by a "settlor," who appoints a trustee to manage certain property for the benefit of ascertainable beneficiaries.[70] Though non-charitable trusts had to have ascertainable beneficiaries at common law,[71] most states now authorize "purpose trusts" created for a declared purpose, even without specific beneficiaries.[72] Anthropic claims the purpose of the Long-Term Benefit Trust "is the same as that of Anthropic," that is, responsibly developing AI for the benefit of humanity.[73] The Anthropic board appointed five initial trustees,[74] who are largely aligned with the effective altruism movement.[75] Every trustee serves a one-year term, and the trustees themselves elect each other.[76] At the time of this writing, two trustees had stepped down without being replaced.[77]

The Long-Term Benefit Trust is intended to police the Anthropic board, but who will police the Trust? In a typical common law trust, the beneficiaries are the principal parties with standing to enforce the terms of the trust and the fiduciary duties of the trustee.[78] Because purpose trusts have no beneficiaries, the settlor must appoint someone who can enforce the terms of the trust (or lacking that, the court will appoint one).[79] Although the Long-Term Benefit Trust agreement is not publicly available, Anthropic's legal advisors assert that it authorizes suits "by the company and by groups of the company's stockholders who have held a sufficient percentage of the company's equity for a sufficient period of time."[80] On its face, this enforcement mechanism is curious because the Long-Term Benefit Trust is supposed to be a check on irresponsible, profit-driven development of AI. While shareholders might sue disloyal trustees who harm the company, presumably they will not sue lax trustees who permit unsafe but profitable strategies.

2. OpenAI. — OpenAI's original structure focused directly on mitigating profit-seeking behavior. The company began as a tax-exempt nonprofit, operating on donations.[81] As with a for-profit corporation,

---

69 *See The Long-Term Benefit Trust*, *supra* note 56; Morley et al., *supra* note 67.

70 Richard C. Ausness, *Non-Charitable Purpose Trusts: Past, Present, and Future*, 51 REAL PROP. TR. & EST. L.J. 321, 324 (2016).

71 Max M. Schanzenbach & Robert H. Sitkoff, *Reconciling Fiduciary Duty and Social Conscience: The Law and Economics of ESG Investing by a Trustee*, 72 STAN. L. REV. 381, 415 (2020).

72 Adam J. Hirsch, *Delaware Unifies the Law of Charitable and Noncharitable Purpose Trusts*, EST. PLAN., Nov. 2009, at 2–4.

73 *The Long-Term Benefit Trust*, *supra* note 56.

74 *Id.*

75 *See* Matthews, *supra* note 64.

76 *The Long-Term Benefit Trust*, *supra* note 56.

77 *Id.*

78 RESTATEMENT (THIRD) OF TRUSTS § 94(1) & cmt. b (AM. L. INST. 2012).

79 *See* Hirsch, *supra* note 72, at 3–4; DEL. CODE ANN. tit. 12, § 3556(c) (2024).

80 Morley et al., *supra* note 67.

81 *See Our Structure*, *supra* note 55.

the board of directors of a nonprofit corporation owes fiduciary duties of care and loyalty.[82] A nonprofit corporation lacks shareholders, so the board "owe[s its] duties to the purposes of the charity."[83] According to OpenAI's certificate of incorporation, the company's purpose is "to provide funding for research, development and distribution of technology related to artificial intelligence. The resulting technology will benefit the public and the corporation will seek to open source technology for the public benefit when applicable."[84] For nonprofit corporations like OpenAI, typically "only directors and the [state] attorney general have standing to sue" to enforce fiduciary duties.[85]

Needing more capital to fund its research, OpenAI created a complicated scheme of new entities in order to raise equity. The nonprofit OpenAI, Inc. remains the top-level entity and its board of directors continues to oversee the entire organization.[86] Several for-profit subsidiaries were created to raise money for the company by selling equity to investors.[87] Employees moved from the nonprofit to one of these entities and received equity as well.[88]

This intricate web of entities helps OpenAI, Inc. preserve its nonprofit, tax-exempt status. The organization's contracts inform investors and employees of its nonprofit mission,[89] and the website advises investors "to view any investment . . . in the spirit of a donation."[90] Investors, including Microsoft, have agreed to cap profits at up to one hundred times their investment.[91] And the partnership agreement for one

---

[82] RESTATEMENT OF CHARITABLE NONPROFIT ORGS. §§ 2.02–.03 (AM. L. INST. 2021).

[83] *Id.* § 2.02 cmt. a.

[84] OpenAI, Inc., Certificate of Incorporation of a Non-Stock Corporation 1 (Dec. 8, 2015).

[85] James J. Fishman, *The Development of Nonprofit Corporation Law and an Agenda for Reform*, 34 EMORY L.J. 617, 677 n.300 (1985). Some states might allow donors to sue to enforce fiduciary duties, but the law of Delaware — where OpenAI, Inc. is incorporated — does not. *See* Samuel D. Brunson, *Musk, OpenAI, and the Internal Affairs Doctrine*, NONPROFIT L. PROF BLOG (Mar. 6, 2024), https://lawprofessors.typepad.com/nonprofit/2024/03/musk-openai-and-the-internal-affairs-doctrine.html [https://perma.cc/5SPS-BXPP]; OpenAI, Inc., *supra* note 84, at 1; Wier v. Howard Hughes Med. Inst., 407 A.2d 1051, 1056–57 (Del. Ch. 1979).

[86] *Our Structure*, *supra* note 55. The corporate structure diagrammed on the website is slightly out of date. *See* Defendant OpenAI's Response to Order to Show Cause Regarding Diversity Jurisdiction ¶¶ 10–14, Walters v. OpenAI, L.L.C., No. 23-cv-03122 (N.D. Ga. Oct. 25, 2023), ECF No. 22.

[87] *See Our Structure*, *supra* note 55.

[88] *Id.*

[89] Greg Brockman & Ilya Sutskever, *OpenAI LP*, OPENAI (Mar. 11, 2019), https://openai.com/index/openai-lp [https://perma.cc/8JLC-CES4].

[90] *Our Structure*, *supra* note 55.

[91] *Id.*; Brockman & Sutskever, *supra* note 89. One hundred times Microsoft's investment of $13 billion would be $1.3 trillion, so OpenAI would need to grow to become one of the largest companies in the world for the cap to limit investors' profits. *See* Tim Bradshaw et al., *How Microsoft's Multibillion-Dollar Alliance with OpenAI Really Works*, FIN. TIMES (Dec. 15, 2023), https://www.ft.com/content/458b162d-c97a-4464-8afc-72d65afb28ed [https://perma.cc/T89E-BPYK]; Daniel Liberto, *Biggest Companies in the World by Market Cap*, INVESTOPEDIA (Oct 16, 2024), https://www.investopedia.com/biggest-companies-in-the-world-by-market-cap-5212784 [https://perma.cc/5FYH-PJ3G].

subsidiary LP — since converted into an LLC[92] — "requir[ed] the partnership 'to give priority to exempt purposes over maximizing profits for the other participants.'"[93] Moreover, OpenAI promises that commercial and intellectual property licenses for artificial general intelligence (AGI) — the stage where AI "outperforms humans at most economically valuable work" — will not benefit investors, but rather "the [n]onprofit and the rest of humanity."[94]

3. Antecedent Scholarly Reactions to OpenAI's and Anthropic's Tools. — While OpenAI and Anthropic are experimenting with new governance structures, elements of these structures have received scholarly treatment in the past. Nonprofit-owned (or nonprofit-controlled) firms are hugely understudied in the modern American economy, likely because there have been very few players since the 1970s.[95] They are more common in many European countries, but still remain an undertheorized facet of corporate law.[96] The research that has been done on these corporations has found them to be relatively successful from a profit standpoint,[97] and they appear to be socially valuable. A major benefit suggested by scholars is foundation-owned firms' heightened ability to either stem short-term profit motives or preserve a commitment to the foundation's mission.[98] But views on the sustainability of such structures are generally mixed. On the one hand, the data suggest that foundation-owned firm structures that are able to keep some separation between foundation directors and for-profit management help to ensure that the management is not coopted by for-profit motives.[99] At Anthropic, trustees — who elect directors — only serve one-year terms, so such loyalty may not be able to attach.[100] On the other hand, studies have also suggested that foundation-owned firms flourish when foundation directors identify strongly with their role as "virtual owners" of the

---

92 *See* Complaint ¶ 10, Musk v. Altman, No. 24-cv-04722 (N.D. Cal. Aug. 5, 2024), ECF No. 1.

93 Ellen P. Aprill et al., *Board Control of a Charity's Subsidiaries: The Saga of OpenAI*, 182 TAX NOTES FED. 289, 292 (2024) (quoting OpenAI, Inc., IRS Form 990: Return of Organization Exempt from Income Tax, Schedule O (OMB No. 1545-0047) (2021)).

94 *Our Structure*, *supra* note 55.

95 *See* Henry Hansmann & Steen Thomsen, *The Governance of Foundation-Owned Firms*, 13 J. LEGAL ANALYSIS 172, 174, 178 (2021).

96 Steen Thomsen & Nikolaos Kavadis, *Enterprise Foundations: Law, Taxation, Governance, and Performance*, 6 ANNALS CORP. GOVERNANCE 227, 316 (2022). *But see* Frederik Hovmark Pedersen, Ownership at OpenAI: From the Perspective of Enterprise Foundation Governance 7–8 (Apr. 15, 2024) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4795279 [https://perma.cc/7NPL-9Y7D].

97 *See* Hansmann & Thomsen, *supra* note 95, at 174.

98 *See id.* at 224–25.

99 *See id.* at 187–88, 206, 220–21.

100 *But see* Alex Kantrowitz, *Oh, Good, OpenAI's Biggest Rival Has a Weird Structure Too*, SLATE (Dec. 2, 2023, 10:00 AM) https://slate.com/technology/2023/12/anthropic-openai-board-trust-effective-altruism.html [https://perma.cc/SH4B-F437] ("Board members will have responsibilities to shareholders, but they won't easily forget those who nominated them and why they did it.").

for-profit entity.[101] This phenomenon is unlikely to appear in Anthropic given the one-year terms and the departure of two trustees already.

Scholars have also expressed mixed views on PBCs.[102] A PBC charter allows the board of directors to balance the interests of stakeholders and shareholders without opening themselves up to duty of loyalty claims.[103] Furthermore, in Delaware, decisions on how best to balance stakeholder and shareholder interests are subject to the business judgment rule[104] — a bedrock doctrine requiring that courts not second-guess business decisions made with care and without conflicts of interest.[105]

Thus, while there has been very little public litigation against directors for purpose-related claims,[106] as a statutory matter, the use of the public benefit corporation charter can largely shield directors from liability for business decisions related to how they treat stakeholders and shareholders. Its efficacy has garnered skepticism from the academic community: Given that no litigation rights are granted to stakeholders and the litigation rights afforded to shareholders for these claims are weakened by substantial deference to directors, scholars doubt the ability of public benefit corporation statutes to hold directors accountable if they choose to sell out the stakeholders.[107] Scholars also highlight the near-redundancy of PBCs; if directors choose to take stakeholders into account, they are largely protected by the business judgment rule even as an ordinary corporation.[108]

Scholarly reactions to OpenAI's and Anthropic's multilayered structures are still forthcoming, but corporate intrigue waits for no man: Dramatic governance-related developments occurred at OpenAI shortly after the firm became a household name.

---

101 Hansmann & Thomsen, *supra* note 95, at 196.

102 *See, e.g.*, Jill E. Fisch & Steven Davidoff Solomon, *The "Value" of a Public Benefit Corporation*, *in* RESEARCH HANDBOOK ON CORPORATE PURPOSE AND PERSONHOOD 68, 69 (Elizabeth Pollman & Robert B. Thompson eds., 2021); Jens Dammann, *Publicly Traded Public Benefit Corporations: An Empirical Investigation,* 29 STAN. J.L. BUS. & FIN. 265, 273 (2024); Ann Lipton, *Benefit Corporations Go Public*, BUS. L. PROF BLOG (July 18, 2020), https://lawprofessors.typepad.com/business_law/2020/07/benefit-corporations-go-public.html [https://perma.cc/4Z9H-X3CG].

103 *See* Fisch & Solomon, *supra* note 102, at 76–77.

104 *See id.* at 77.

105 *See* Aronson v. Lewis, 473 A.2d 805, 811–13 (Del. 1984).

106 *See* Amy L. Simmerman et al., *Converting to a Delaware Public Benefit Corporation: Lessons from Experience*, HARV. L. SCH. F. ON CORP. GOVERNANCE (Feb. 18, 2022), https://corpgov.law.harvard.edu/2022/02/18/converting-to-a-delaware-public-benefit-corporation-lessons-from-experience [https://perma.cc/SY2N-4HMY].

107 *See, e.g.*, Fisch & Solomon, *supra* note 102, at 74–75, 82–84; Dammann, *supra* note 102, at 279–80; Lipton, *supra* note 102; *see also* DEL. CODE ANN. tit. 8, § 367 (2024) (limiting standing for suits to enforce a public benefit corporation's purpose to shareholders who own at least 2% of shares or $2 million of shares).

108 *See supra* notes 104–05 and accompanying text.

C. Shareholders, Stakeholders, and Shakeups

In November 2023, OpenAI made headlines when its board fired CEO Sam Altman and rehired him a few days later after employees threatened to quit en masse and accept jobs at Microsoft.[109] When the dust settled, most of the directors who had fired Altman were gone.[110] In keeping with OpenAI's commercial pivot, the old, AI safety–focused directors were replaced with directors mostly drawn from the heights of government and the tech sector.[111] One ex–board member framed the firing in terms of the board's duties: "Our goal in firing Sam was to strengthen OpenAI and make it more able to achieve its mission . . . . [T]he nonprofit mission — to ensure AGI benefits all of humanity — comes first."[112] In 2024, Elon Musk — an early funder of the OpenAI nonprofit who now happens to own a rival AI startup himself[113] — sued Altman and OpenAI for allegedly "betray[ing]"[114] the company's nonprofit mission by partnering with Microsoft to monetize OpenAI's technology.[115]

Recently, the company has declared it will convert its for-profit entity to a PBC, much like Anthropic, and will grant the nonprofit entity a "significant" stake.[116] This move will give the for-profit company wide latitude in determining how best to balance its mission against its profit. Speaking in generalities, OpenAI and Anthropic seemed to adopt a stakeholderist approach, driven by the fear that shareholders could derail their missions to develop AI safely. But, despite these attempts, OpenAI has undergone tumultuous changes that many have described

---

[109] Cade Metz et al., *Sam Altman Is Reinstated as OpenAI's Chief Executive*, N.Y. TIMES (Nov. 22, 2023) https://www.nytimes.com/2023/11/22/technology/openai-sam-altman-returns.html [https://perma.cc/2KYA-CX5R].

[110] *Id.*

[111] *See* Cade Metz, *Key Players in OpenAI's Boardroom Drama*, N.Y. TIMES (Dec. 9, 2023), https://www.nytimes.com/2023/12/09/technology/openai-boardroom-key-players.html [https://perma.cc/XN8A-MNBX]; *OpenAI Announces New Members to Board of Directors*, OPENAI (Mar. 8, 2024), https://openai.com/index/openai-announces-new-members-to-board-of-directors [https://perma.cc/7F25-BP4Z].

[112] Meghan Bobrowsky & Deepa Seetharaman, *The OpenAI Board Member Who Clashed with Sam Altman Shares Her Side*, WALL ST. J. (Dec. 7, 2023, 4:39 PM), https://www.wsj. com/tech/ai/helen-toner-openai-board-2e4031ef [https://perma.cc/R696-LRZ5] (quoting Helen Toner). *But see* Bret Taylor & Larry Summers, *OpenAI Board Members Respond to a Warning by Former Members*, THE ECONOMIST (May 30, 2024), https://www.economist.com/by-invitation/2024/05/30/openai-board-members-respond-to-a-warning-by-former-members [https://perma.cc/ZH9U-KFJ4].

[113] Anna Tong et al., *OpenAI Asks Investors to Avoid Five AI Startups Including Sutskever's SSI, Sources Say*, REUTERS (Oct. 2, 2024, 6:22 PM), https://www.reuters.com/technology/openai-tells-investor-not-invest-five-ai-startups-including-sutskevers-ssi-2024-10-02 [https://perma.cc/38F3-DMSU].

[114] Complaint, *supra* note 92, ¶ 2.

[115] *See id.* ¶¶ 1–3.

[116] *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/8BVD-QCBF].

as selling out its social mission.[117] These developments invite the question of how to update theories of amoral drift for private, capital-intensive startups and for AI companies specifically.

This section reimagines the process of amoral drift in the context of AI companies — nascent, private capital–backed firms with substantial capital requirements. This section first proposes that by giving equity to groups of critical stakeholders, OpenAI's and Anthropic's novel corporate governance techniques ended up creating an even more dangerous faction than shareholders: "superstakeholders" (that is, employees and suppliers with immense profit interests). This section then describes the chaotic events at OpenAI in terms of this paradigm.

1. Amoral Drift. — AI companies diverge substantially from the hypothetical company central to Hart and Zingales's model of amoral drift. Hart and Zingales's amoral drift thought experiment contemplates a founder's ability to preserve their company's prosocial mission after the company goes public.[118] From the pre-ESG era through the present, "corporate purpose" literature has largely been preoccupied with controlling dispersed shareholder bases, a threat mainly associated with public or soon-to-be public firms.[119] This is not without good reason: Before the recent private capital wave, listing on a public exchange was the primary way companies could gain access to large amounts of capital to invest in further growth.[120] Thus, most discussions of growing companies were geared toward an eventual initial public offering (IPO).[121] When scholars have described how shareholder pressure robs corporations of their social orientation, private or closely held companies are often used as theoretical foils to the public firms at the center of the discussion.[122] As a result, the process in which private companies struggle to preserve a prosocial mission remains undertheorized.

Staying private[123] means that AI startups like OpenAI and Anthropic have strong profit incentives that have not been realized

---

117 *See, e.g.*, Sigal Samuel, *OpenAI as We Knew It Is Dead*, VOX (Sept. 26, 2024, 5:01 PM), https://www.vox.com/future-perfect/374275/openai-just-sold-you-out [https://perma.cc/JM44-EN2Q].

118 *See supra* section A, pp. 1636–39.

119 *See, e.g.*, Elhauge, *supra* note 19, at 742; Dana Brakman Reiser & Steven A. Dean, *Hunting Stag with FLY Paper: A Hybrid Financial Instrument for Social Enterprise*, 54 B.C. L. REV. 1495, 1504–05 (2013) [hereinafter Reiser & Dean, *Hunting Stag*]; Dana Brakman Reiser & Steven A. Dean, *Financing the Benefit Corporation*, 40 SEATTLE U. L. REV. 793, 795–96 (2017).

120 *See* Elisabeth de Fontenay, *The Deregulation of Private Capital and the Decline of the Public Company*, 68 HASTINGS L.J. 445, 448 (2017).

121 *See id.* at 459–60.

122 *See, e.g.*, Aguirre, *supra* note 50, at 2111 ("[I]t is perhaps not surprising that one of the most prominent large-scale examples of a company that articulates objectives beyond profit, Patagonia, is privately held.").

123 *See generally* Daria Davydova et al., *Why Do Startups Become Unicorns Instead of Going Public?* (Eur. Corp. Governance Inst., Working Paper No. 857/2022, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4899183 [https://perma.cc/6V4M-LQES] (finding that most "unicorns,"

(through an IPO, acquisition, or some other liquidity event), but they retain the structure of closely held private companies — there are no diffuse shareholder bases to consider. Thus, the major players consist of the founders, early investors, and stakeholders.

Scholars have noted that corporate governance has entered the age of the strong stakeholder.[124] Through social media, consumers have overcome collective action problems to publicly shame companies and CEOs.[125] Employees have engaged in strike activity not seen since prior to the 2020 pandemic.[126] The "strong stakeholder" dynamics are especially salient for AI companies, which have furnished employees and tech suppliers with valuable equity stakes. Thus, AI companies have seen the rise of not just the strong stakeholder but also the "superstakeholder" — a stakeholder, supercharged by equity stakes with staggering upside potential, whose interests have subsumed those of the shareholder. Employees and tech suppliers are distinguishable from traditional shareholders in that they are not just nominal owners waiting to benefit from growth; they are capable of crippling and even destroying a company because their presence is essential. OpenAI and Anthropic have attempted to tackle what they perceived to be the most probable threat to their mission — the profit motives of overzealous shareholders — by structurally minimizing the avenues for shareholders to mobilize.[127] But in doing so, they have moved the profit-based compensation (equity) to new parties — stakeholders — and ended up shoring up defenses against the virtually nonexistent threat of traditional shareholders. The result for these AI companies is that critical stakeholders motivated by profit are not counterbalanced.[128] Thus, unconstrained stakeholders are free to instigate amoral drift.

The facts of the AI-startup reality suggest that prosocial founders' foci ought to be not on shareholders but on major stakeholders. However, OpenAI's and Anthropic's current tools to lock in their founders' prosocial visions are structurally geared toward defending against shareholder-led amoral drift and are less powerful against stakeholder-led amoral drift. The following sections try to unravel the mystery of OpenAI's corporate drama, describe the wide range of profit-focused

---

privately-held companies valued at over $1 billion, are staying private for long periods after they are able to go public, *id.* at 1); Hayden Field & MacKenzie Sigalos, *AI Craze Is Distorting VC Market, As Tech Giants Like Microsoft and Amazon Pour In Billions of Dollars*, CNBC (Sept. 6, 2024, 10:47 AM), https://www.cnbc.com/2024/09/06/ai-craze-getting-funded-by-tech-giants-distorting-traditional-vcs.html [https://perma.cc/G8DK-GALD].

124 *See, e.g.*, Michal Barzuza et al., *The Millennial Corporation: Strong Stakeholders, Weak Managers*, 28 STAN. J.L. BUS. & FIN. 255, 259, 284 (2023).

125 *Id.* at 282–84.

126 Margaret Poydock & Jennifer Sherer, *Major Strike Activity Increased by 280% in 2023*, ECON. POL'Y INST. (Feb. 21, 2024), https://www.epi.org/publication/major-strike-activity-in-2023 [https://perma.cc/25FW-X3ME].

127 *See supra* sections B.1–2, pp. 1640–43.

128 *See* Weiss, *supra* note 16 (manuscript at 6).

actors involved with OpenAI and Anthropic, and speculate about key drivers of stakeholder-led amoral drift.

2. What Happened at OpenAI? — Outside observers may never fully understand what happened at OpenAI when Sam Altman was abruptly fired and subsequently reinstated. But one can imagine the following: Amoral drift was initiated by stakeholders, and when the nonprofit board tried to assert its own power, it found itself weakened in relation to employees and Big Tech. Compromised in comparison to powerful superstakeholders, the board had little choice but to acquiesce or risk the destruction of the entire enterprise — social or otherwise.

(a) Founders. — The shuffle at OpenAI invites a sobering question: Were the OpenAI founders ever really committed to safe AI development in the first place? The narratives surrounding the firm's inception and development suggest that there may be no easy answer. The founders at OpenAI and Anthropic demonstrate the complicated mix of prosocial and profit-focused aims animating AI company founders. Dynamics at both firms are largely aligned with what Hart and Zingales theorize: Founders are not always strictly in favor of or against social aims.[129] OpenAI's founders and initial funders — including Elon Musk, Sam Altman, Peter Thiel, Reid Hoffman, and Jessica Livingston — displayed at the outset an interest in developing AI without a focus on profit.[130] OpenAI was initially formed as a research foundation, in part so that the founders could have complete control over AI development without being constrained by a duty to maximize profit.[131] But over time, and as Musk stepped away, the founders understood they could not competitively develop the technology without substantially more capital, so they began incorporating for-profit elements into their structure.[132] Anthropic, founded by former OpenAI employees, also expressed a commitment to safety at the time of the firm's creation. It reaffirmed that commitment when it reorganized the structure of the company in 2023.[133]

(b) VCs. — Venture capitalists (VCs) are the actors that most resemble traditional shareholders. VCs, initial investors in early-stage companies that stand to gain outsized returns, generally focus on profits and typically invest through preferred stock.[134] Preferred stockholders have a more senior claim to any payouts on the sale or dissolution of the company relative to common stockholders.[135] VC investors tend to have specific expertise to share; as a result, they often take board seats, have

---

129 *See* Hart & Zingales, *supra* note 20, at 252–54.

130 *Introducing OpenAI*, OPENAI (Dec. 11, 2015), https://openai.com/index/introducing-openai [https://perma.cc/BK7Q-B95K].

131 *Id.* ("As a non-profit, our aim is to build value for everyone rather than shareholders.").

132 *See Why OpenAI's Structure Must Evolve to Advance Our Mission*, *supra* note 116.

133 *The Long-Term Benefit Trust*, *supra* note 56.

134 *See* Pollman, *supra* note 64, at 172–73.

135 *Id.* at 173 & n.91.

very close relationships with management, and can exercise some level of control over corporate decisions.[136] While VCs can play a monitoring role using their expertise to oversee the reasoned development of a new company, the VC model of investing broadly in the hopes of "one or two 'home runs'"[137] gives VCs an incentive to encourage high-risk strategies.[138]

Prior to the creation of the for-profit entity, OpenAI relied on funding from Elon Musk and primarily encouraged other investment "in the spirit of a donation."[139] It therefore did not engage in traditional fundraising targeting VC investors. Prominent VC firms including Thrive Capital and Andreessen Horowitz invested in OpenAI by purchasing shares owned by employees.[140]

(c) Employees. — When OpenAI's nonprofit board fired Sam Altman, employees expressed surprise and threatened to resign in droves.[141] Employees are natural casualties of a corporation's exclusive focus on profit because suppressing wages helps lower expenses and boost profits. But AI company employees do not fit this mold. Startups, including AI companies, often compensate employees with potentially very valuable equity.[142] To provide a striking example: In February 2024, OpenAI closed a sale that allowed employees to sell their equity to Thrive Capital.[143] The pricing of the equity produced a company valuation of $80 billion, up from $29 billion the previous year, meaning that the value of employees' equity had increased by nearly 300% since the previous sale.[144] As a point of comparison, the S&P 500 produced a total three-year return of 7.3% with a standard deviation of 17.4% as of January 2, 2025.[145] The ability of equity to turn employees into profit-focused capitalists has been widely studied,[146] but AI companies are unique in that the risk of depressed wages and layoffs is not a countervailing

---

136 *Id.* at 173; Darian M. Ibrahim, *Corporate Venture Capital*, 24 U. PA. J. BUS. L. 209, 215–16 (2021).

137 Brian J. Broughman & Matthew T. Wansley, *Risk-Seeking Governance*, 76 VAND. L. REV. 1299, 1303 (2023).

138 *See id.* at 1303–04.

139 *Our Structure*, *supra* note 55; *see* Complaint, *supra* note 92, ¶ 1.

140 Cade Metz & Tripp Mickle, *OpenAI Completes Deal that Values the Company at $80 Billion*, N.Y. TIMES (Feb. 16, 2024), https://www.nytimes.com/2024/02/16/technology/openai-artificial-intelligence-deal-valuation.html [https://perma.cc/54WP-HNY8].

141 *See* Duffy & Madhok, *supra* note 18.

142 Weiss, *supra* note 16 (manuscript at 20–21, 22–23).

143 Metz & Mickle, *supra* note 140.

144 *Id.*

145 *See S&P 500 PR*, MORNINGSTAR (Jan. 2, 2025, 5:12 PM), https://www.morningstar.com/indexes/spi/spx/risk [https://perma.cc/U8XZ-AU6J].

146 *See, e.g.*, Martin Gelter, *The Pension System and the Rise of Shareholder Primacy*, 43 SETON HALL L. REV. 909, 911–12 (2013); Saeyoung Chang, *Employee Stock Ownership Plans and Shareholder Wealth: An Empirical Investigation*, 19 FIN. MGMT. 48, 48 (1990).

consideration for AI developers.[147] OpenAI employees who threatened to leave would not be out in the job market for long; the market for skilled AI developers is booming.[148]

Despite being compensated with equity that could make them millionaires overnight, employees of OpenAI have previously had relatively few opportunities to cash out. Recently, OpenAI has expressed a move toward conducting more frequent equity sale opportunities in an attempt to appease current and former employees.[149] This shift supports a general sentiment of anxiety regarding employee equity compensation. It is not difficult to imagine the anxieties that would arise amongst OpenAI employees upon learning of Altman's removal, especially given external investors were pulling out of employee liquidity transactions and considering revaluing the company's equity at zero.[150] Employees as profit-driven superstakeholders were able to weaponize their immense leverage against the nonprofit board to bring back Altman.

(d) Large Tech Companies. — Microsoft played a key role in the chaos after the OpenAI board fired Sam Altman, "assur[ing] . . . positions for all OpenAI employees"[151] and "[p]laying a central role in negotiations" to reinstate Altman as CEO.[152] Here, Microsoft represents another hybrid stakeholder — large tech companies. These companies play many roles. As investors, Big Tech is looking to benefit from the growth in firm value.[153] The AI market has been incredibly competitive, and some large companies have decided to coopt burgeoning startups rather than focus exclusively on developing an AI arm in-

---

147 *See* Katherine Bindley, *The Fight for AI Talent: Pay Million-Dollar Packages and Buy Whole Teams*, WALL ST. J. (Mar. 27, 2024, 2:20 PM), https://www.wsj.com/tech/ai/the-fight-for-ai-talent-pay-million-dollar-packages-and-buy-whole-teams-c370de2b [https://perma.cc/SHP2-UBSN].

148 *Id.*

149 *See* Kaili Killpack, *OpenAI's Big Holiday Payday: How 400 Employees Could Walk Away with $10 Million Each*, BENZINGA (Dec. 28, 2024, 8:00 AM), https://www.benzinga.com/startups/24/12/42713616/openais-big-holiday-payday-how-400-employees-could-walk-away-with-10-million-each [https://perma.cc/Y7J5-MYE6].

150 *See* Ashlee Vance et al., *Nearly All of OpenAI Staff Threaten to Go to Microsoft If Board Doesn't Quit*, BLOOMBERG (Nov. 20, 2023, 4:06 PM), https://www.bloomberg.com/news/articles/2023-11-20/openai-staff-threaten-to-go-to-microsoft-if-board-doesn-t-quit [https://perma.cc/TLF2-TBDJ].

151 *Id.*

152 Emily Chang et al., *OpenAI Leaders' Efforts to Bring Back Altman Reach Impasse over Board Role*, BLOOMBERG (Nov. 19, 2023, 5:21 PM), https://www.bloomberg.com/news/articles/2023-11-19/openai-negotiations-to-reinstate-altman-hit-snag-over-board-role [https://perma.cc/27R2-CFPT].

153 *See* Anat Alon-Beck, *Alternative Venture Capital: The New Unicorn Investors*, 87 TENN. L. REV. 983, 1017–18 (2020); Joseph A. McCahery et al., *Corporate Venture Capital: From Venturing to Partnering*, *in* THE OXFORD HANDBOOK OF VENTURE CAPITAL 211, 218 (Douglas Cumming ed., 2012). To note, some observers view corporate venture capital as a primarily strategic endeavor, *see* Weiss, *supra* note 16 (manuscript at 6), while others suggest corporate venture capital has a mix of strategic and financial objectives, *see* Ibrahim, *supra* note 136, at 224.

house.[154] Many see Microsoft as the de facto owner of OpenAI,[155] and Google and Amazon have both poured staggering amounts of capital into Anthropic.[156] The markets have been receptive to Big Tech's co-optation of AI startups.[157] On the other hand, Samsung, which failed to either develop robust AI capability in-house or partner with an AI startup, declined in value by over $120 billion due to investor concerns that it was losing the AI race.[158] Microsoft, too, suffered a "notable loss" in the value of its stock after news of Altman's firing initially broke.[159] As such, Big Tech's own stock prices and market positions are still tied to their presence in the AI sector and their relationships with AI companies. These large tech companies also act as suppliers, licensors, and business partners. For example, a substantial portion of Microsoft's multibillion dollar investment into OpenAI is believed to consist of

---

[154] *See* Matt O'Brien & Sarah Parvini, *US Senators Call Out Big Tech's New Approach to Poaching Talent, Products from Smaller AI Startups*, ASSOCIATED PRESS (July 12, 2024, 3:09 PM), https://apnews.com/article/ai-artificial-intelligence-acquihires-amazon-adept-wyden-fa3cd0502a757e5a9ccb83d00bf9ad55 [https://perma.cc/ZAQ3-BEBM]; Deirdre Bosa & Jasmine Wu, *The Sneaky Way Big Tech Is Acquiring AI Unicorns Without Buying the Companies*, CNBC (Aug. 30, 2024, 1:47 PM), https://www.cnbc.com/2024/08/30/how-google-microsoft-and-amazon-are-raiding-ai-startups-for-talent.html [https://perma.cc/X3TR-89K8]. *See also generally* Mark A. Lemley & Matthew T. Wansley, *Coopting Disruption*, 105 B.U. L. REV. (forthcoming 2025), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4713845 [https://perma.cc/RYV5-S4RP] (describing Big Tech's systemic tendencies to initially support tech startups and then stifle competition and ultimately acquire disruptive tech companies).

[155] *See* Matt Levine, Opinion, *Who Owns OpenAI?*, BLOOMBERG (Oct. 21, 2024, 2:52 PM), https://www.bloomberg.com/opinion/articles/2024-10-21/who-owns-openai [https://perma.cc/544Q-KKSB].

[156] *See* Jackie Davalos & Brad Stone, *OpenAI Rival Anthropic Defends Partnerships with Amazon, Google*, BLOOMBERG (May 9, 2024, 1:40 PM), https://www.bloomberg.com/news/articles/2024-05-09/openai-rival-anthropic-defends-partnerships-with-amazon-google [https://perma.cc/62TP-374D]. These investments have not gone without scrutiny from antitrust regulators. Previously, Microsoft was required to relinquish its observer seat on OpenAI's board due to antitrust concerns. *See* Jyoti Mann, *Microsoft and Apple May Be Playing the Long Game by Ditching OpenAI Board Roles*, BUS. INSIDER (July 10, 2024, 7:21 AM), https://www.businessinsider.com/microsoft-apple-ditch-openai-board-observer-seats-regulators-2024-7 [https://perma.cc/ZVS6-TAXB]. In October 2024, the United Kingdom's Competition and Markets Authority announced that it was investigating Google's $2 billion investment into Anthropic. *See* Mauro Orru & Ian Walker, *Google's $2 Billion Anthropic Investment Faces U.K. Antitrust Scrutiny*, WALL ST. J. (Oct. 24, 2024, 11:56 AM), https://www.wsj.com/tech/ai/googles-pact-with-anthropic-probed-by-u-k-regulator-3215637c [https://perma.cc/U3EQ-77XX].

[157] *See, e.g.*, Jordan Novet, *Microsoft Closes at All-Time High on Fresh OpenAI-Related Optimism*, CNBC (Nov. 7, 2023, 10:53 PM), https://www.cnbc.com/2023/11/07/microsoft-closes-at-all-time-high-on-fresh-openai-related-optimism.html [https://perma.cc/6RFN-BS4X]; Emma Cosgrove & Ben Bergman, *Amazon Makes Massive Downpayment on Dethroning Nvidia*, BUS. INSIDER (Nov. 22, 2024, 2:09 PM), https://www.businessinsider.com/amazon-tries-again-anthropic-ai-chips-trainium-nvidia-2024-11 [https://perma.cc/HC8U-SKHN] (noting Nvidia stock's drop after Amazon announced additional investment in Anthropic).

[158] *See* Arjun Kharpal, *How Samsung Fell Behind in the AI Boom Leading to a $126 Billion Wipeout*, CNBC (Nov. 8, 2024, 4:27 PM), https://www.cnbc.com/2024/11/08/how-samsung-fell-behind-in-the-ai-boom-behind-rival-sk-hynix.html [https://perma.cc/65HL-NJ5E].

[159] Rachyl Jones, *Satya Nadella Added $63 Billion in Market Value for Microsoft with a "Poker Move for the Ages*," FORTUNE (Nov. 20, 2023, 1:56 PM), https://fortune.com/2023/11/20/satya-nadella-63-billion-market-value-microsoft-openai-poker-move [https://perma.cc/9T7W-AG4X].

cloud computing credits.[160] AI requires large amounts of "compute,"[161] fostering a symbiotic relationship between AI startups and Big Tech. In fact, Microsoft's services give it so much leverage that Altman once told an interviewer that "if Microsoft were to cut [OpenAI] off from its servers, [OpenAI's] work would be effectively paralyzed."[162] Thus, employees and Big Tech working in concert threatened the very existence of OpenAI.

3. Superstakeholder-Led Amoral Drift. — While this Chapter can only speculate, it is possible that OpenAI's nonprofit board miscalculated, or misread, the power of the stakeholders. OpenAI's nonprofit board had no shareholders and no fiduciary duty to boost profits. Furthermore, OpenAI's nonprofit board was quite far removed from the day-to-day experience of employees, as well as the for-profit entity's reliance on Microsoft for investment and compute. It's possible that absent ties to a traditional shareholder base, the nonprofit board could not immediately sense the temperature of employees and Microsoft on firing Sam Altman. The current chair of OpenAI's nonprofit board, Bret Taylor, has made remarks consistent with this dynamic, suggesting the existence of a disconnect even prior to OpenAI's board shuffle: After describing his role as one of "governance" rather than "day-to-day operations,"[163] he contrasted OpenAI's aim of "building artificial general intelligence" with the aims of his other company, Sierra, which he described as "creating a product for enterprises."[164] Although Taylor may view his role as focused on governing research into AGI rather than overseeing a commercial, product-oriented enterprise, OpenAI's equity-compensated employees and investors as well as customers may not feel the same.

If this process of amoral drift did occur, it is unlikely that organizing as a PBC would have prevented it. Additionally, the nonprofit-

---

160 *See* Reed Albergotti, *OpenAI Has Received Just a Fraction of Microsoft's $10 Billion Investment*, SEMAFOR (Nov. 18, 2023, 3:28 PM), https://www.semafor.com/article/11/18/2023/openai-has-received-just-a-fraction-of-microsofts-10-billion-investment [https://perma.cc/MQC6-XWHP].

161 *See* JAI VIPRA & SARAH MYERS WEST, COMPUTATIONAL POWER AND AI 5 (2023), https://ainowinstitute.org/wp-content/uploads/2023/09/AI-Now_Computational-Power-an-AI.pdf [https://perma.cc/CG9P-HW66]; *see also Microsoft CEO Satya Nadella on the OpenAI Debacle*, PIVOT (Nov. 20, 2023), https://podcasts.apple.com/us/podcast/microsoft-ceo-satya-nadella-on-the-openai-debacle/id1073226719?i=1000635493753 [https://perma.cc/25DS-VNGP] ("[T]here is no OpenAI without . . . Microsoft leaning in . . . a deep way to partner with this company on their mission," *id.* at 3:56 (statement of Satya Nadella, CEO of Microsoft)).

162 Max Chafkin & Dina Bass, *Microsoft's Sudden AI Dominance Is Scrambling Silicon Valley's Power Structure*, BLOOMBERG (June 15, 2023, 10:49 AM), https://www.bloomberg.com/news/features/2023-06-15/microsoft-prepares-to-cash-in-on-openai-partnership-with-copilot [https://perma.cc/QZ2R-4EJ2].

163 Kylie Robison, *OpenAI Chair Bret Taylor Says He'll Recuse Himself "Whenever There Is a Potential for Overlap" with His New AI Startup Sierra*, FORTUNE (Feb. 13, 2024, 7:12 PM), https://fortune.com/2024/02/13/openai-chair-bret-taylor-interview-promises-recuse-whenever-potential-overlap-ai-startup-sierra [https://perma.cc/4S8P-Y6J8].

164 *Id.*

controlled structure likely exacerbated the disjuncture between OpenAI's nonprofit board and its stakeholders. AI companies may need to look elsewhere for ways to preserve their social missions in the face of superstakeholders.

D. Amoral Drift: Inevitable or Avoidable?

As the previous sections have discussed, OpenAI's novel corporate governance measures incompletely shielded the board's decisionmaking from profit-oriented superstakeholders. These employees and suppliers (Microsoft) had appetites for risk similar to shareholders because their compensation incorporated a stake in future profits.[165] But the OpenAI employees and Microsoft have far more leverage over the company than typical shareholders because they also provide scarce, mission-critical resources: in the employees' case, highly skilled labor, and in Microsoft's case, compute.[166]

OpenAI compensated its employees and Microsoft with equity out of necessity. Providers of scarce resources can demand higher compensation. But startups generally don't make enough money to afford to pay so much compensation in cash.[167] While loans are an option for more mature companies, banks are generally unwilling to advance funds to startups, whose uncertain business prospects imperil timely repayment.[168] As a result, startups typically compensate employees with equity — that is, a piece of the (hopefully gigantic) profits they will make in the future — because startups have little else to give.[169] Similarly, startups sometimes compensate service providers such as lawyers — and in OpenAI's case, Microsoft — with equity as a form of deferred compensation in place of or in addition to fees.[170]

This phenomenon of equity compensation complicates attempts to use corporate governance to minimize externalities. Employees and other stakeholders with equity compensation will want the company to make more profitable (but potentially riskier) choices. Conversely, entirely salaried employees and supplier-creditors will desire stabler, less risky (but potentially less profitable) choices. After all, a stable company will be a stable employer and a reliable partner for suppliers. Riskier strategies might lead to losses, causing employees to be laid off and

---

165 *See supra* sections C.2.c–d, pp. 1649–52.

166 *See supra* sections C.2.c–d, pp. 1649–52; *see also* Pollman, *supra* note 64, at 193–94.

167 Pollman, *supra* note 64, at 171.

168 *See* Elizabeth Pollman, *Startup Failure*, 73 DUKE L.J. 327, 335–36 (2023). *But see* J. Brad Bernthal, *The Evolution of Entrepreneurial Finance: A New Typology*, 2018 BYU L. REV. 773, 814 fig.1, 822 fig.2 (cataloging innovative debt-like instruments used in startup financings); Darian M. Ibrahim, *Debt as Venture Capital*, 2010 U. ILL. L. REV. 1169, 1173 (exploring phenomenon of debt financing that follows venture capital equity investments).

169 Pollman, *supra* note 64, at 171–72.

170 Sarah Boulden, Student Note, *The Business of Startup Law: Alternative Fee Arrangements and Agency Costs in Entrepreneurial Law*, 11 J. ON TELECOMM. & HIGH TECH. L. 279, 294–95 (2013).

suppliers to be left unpaid. With equity compensation, however, employees and suppliers receive huge potential upside that compensates for the downside risk of losses or bankruptcy. If startups have little else of value but equity, how can startups like OpenAI adequately compensate critical stakeholders without creating misaligned superstakeholders?

One potential answer is equity compensation linked to mitigation of AI risk. Corporations in recent years have tied their executives' pay to measures of ESG performance.[171] For example, CEO compensation has been tied to improving employee diversity and reducing carbon emissions, among other goals.[172] AI startups could similarly attempt to align employees' incentives with the founder's prosocial goals through strategic grants of equity. Performance share units (PSUs) are commonly used instruments that typically "deliver[] a variable number of shares at the end of a three-year performance period."[173] With PSU grants, employees and executives will receive more shares at the end of the performance period if the company achieves AI safety goals and avoids AI accidents or fewer shares if AI threats come to pass.

However, there are several shortcomings with ESG-linked compensation, which could reduce its effectiveness for AI companies. First, AI safety metrics may be hard to measure, so the entity charged with measuring will have outsized influence over employee pay.[174] Typically, the board or the board's compensation committee handles the specifics of executive pay.[175] The boards of Anthropic and (at least initially) OpenAI are structurally geared toward AI safety. So, in theory, these boards could plausibly be trusted to impartially calculate AI safety–linked compensation.[176] However, if the board might drift from the original safety-centric mission (as occurred at OpenAI), it would not be wise to vest discretion over easily manipulable compensation-related AI safety metrics in the board. An external referee in the mold of Anthropic's Long-Term Benefit Trust is likely the most trustworthy entity for ensuring compensation is aligned with prosocial goals.

Second, it may be difficult to square the time horizon of equity compensation with the time horizon of AI risk. On one hand, some manifestations of AI risk are more or less immediately perceptible: for

---

171 Lucian A. Bebchuk & Roberto Tallarita, *The Perils and Questionable Promise of ESG-Based Compensation*, 48 J. CORP. L. 37, 45–46 (2022).

172 *Id.* at 57 tbl.3.

173 David I. Walker, *The Economic (In)significance of Executive Pay ESG Incentives*, 27 STAN. J.L. BUS. & FIN. 318, 328 (2022).

174 *Cf.* Bebchuk & Tallarita, *supra* note 171, at 63, 67 (discussing CEOs' ability to influence their own pay packages, particularly when pay is based on subjective or difficult-to-measure criteria).

175 *See* Walker, *supra* note 173, at 345–46.

176 *But cf.* Walker, *supra* note 173, at 348–49; Adam B. Badawi & Robert Bartlett, *ESG Overperformance? Assessing the Use of ESG Targets in Executive Compensation Plans* 4 (John M. Olin Program in L. & Econ., Working Paper No. 592, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4941016 [https://perma.cc/8M6Q-VWFH] (finding ESG performance goals are set such that executives nearly always attain them).

example, discoveries of AI bias or intellectual property violations.[177] In these cases, it should be relatively easy to adjust compensation to account for internal failures. But some hazards may only surface years down the road. If plausibly dangerous superintelligent AI is decades away, a researcher working on pieces of the puzzle now will not feel constrained by conditions on equity that vests in three years. Extending the vesting period to significantly longer than three years is an infeasible solution because long-delayed compensation would be unattractive to prospective employees.

Creative debt instruments offer another potential answer for compensating stakeholders. Scholars have proposed "corporate social responsibility bonds" (CSR bonds), which do not require repayment if prosocial goals are met,[178] and "flexible low yield paper" (FLY paper), which is low-cost debt that converts into equity if founders abandon the prosocial mission.[179] These instruments may prove useful for prosocial AI startups because they can use the capital to pay non-prosocial stakeholders in cash rather than equity, preventing the creation of superstakeholders. Furthermore, shareholders and profit-motivated stakeholders will have to think twice before pressuring the board to drift from the prosocial mission because it will affect their bottom line. The CSR bonds will reduce the company's profits if the bonds have to be repaid, and the FLY paper will dilute the existing equityholders if it converts to equity.

The main obstacle to issuing these creative types of debt is the limited supply of prosocial capital. AI startups cannot effectively fund themselves with this debt if nobody wants to buy it. Private, closely held companies like startups cannot tap public markets to access the capital of the ordinary, prosocial people contemplated by Hart and Zingales.[180] Startups rely on wealthy individuals and institutional investors for capital, only some of whom will be able and willing to invest prosocially.[181] Indeed, OpenAI has justified its transition to a for-profit structure by pointing to the inadequacy of donations and capped profits for meeting the enormous costs of developing AGI.[182]

Ultimately, the promise of any corporate governance–oriented solution to AI risk is bounded by its reliance on prosocial corporate actors. Employees skeptical of the warnings about AI may choose to work at "dirty" AI companies with higher, less rule-bound compensation. Investors who disbelieve the so-called "AI doomers" may offer their capital to

---

177 *See supra* notes 11–12 and accompanying text.
178 Lund, *supra* note 16, at 1637.
179 *See* Reiser & Dean, *Hunting Stag*, *supra* note 119, at 1497–98.
180 *See* Pollman, *supra* note 64, at 163–65.
181 *See id.* at 167.
182 *See Why OpenAI's Structure Must Evolve to Advance Our Mission*, *supra* note 116.

"dirty" startups offering higher returns instead of clean ones.[183] If citizens think AI does in fact pose serious threats to society, they should not leave it to private ordering to solve the problem.[184]

Conclusion

This Chapter has reimagined "amoral drift" for applicability to the unique landscape of AI companies, which are private, closely held companies with founder staying power — all elements that Hart and Zingales did not have in mind. And yet analysis of these elements suggests that amoral drift may still be inevitable despite efforts to prevent it. Ultimately, by solving for shareholder pressure instead of stakeholder pressure, AI companies may have solved for the wrong variable. Future prosocial innovation in AI corporate governance must thread the needle between raising capital and reducing the influence of profit-motivated actors. Despite the industry's admirable willingness to innovate and experiment, recent experiments do not appear to have met the challenge. Whether AI will prove as risky as its critics predict remains an open question. But the attempts to address concerns through corporate governance have revealed the multifaceted nature of shareholders and stakeholders, confounding efforts to solve amoral drift.

---

[183] *See* Tallarita, *supra* note 20. As this Chapter was finalized, a Musk-led investor group formally offered to purchase OpenAI for $97.4 billion. Jessica Toonkel & Berber Jin, *Elon Musk-Led Group Makes $97.4 Billion Bid for Control of OpenAI*, WALL ST. J. (Feb. 10, 2025, 4:37 PM), https://www.wsj.com/tech/elon-musk-openai-bid-4af12827 [https://perma.cc/27PZ-4F8D]. While the bid was summarily rejected by Altman, *id.*, the offer could complicate the planned spinoff of the for-profit entity. For one, Musk's valuation may raise the baseline for the fair price required to compensate the nonprofit for the spinoff of its for-profit subsidiary. And as this Chapter has detailed, the interests of founders, VCs, employees, and Big Tech may not be straightforward as the groups assess Musk's bid and its implications on the transfer of their previous investments into the new for-profit entity. Musk's inconsistent attitudes toward OpenAI's prosocial orientation, *see* Scott Rosenberg, *Musk Is AI Policy's Giant Question Mark*, AXIOS (Nov. 19, 2024), https://www.axios.com/2024/11/19/elon-musk-trump-ai-regulation-path [https://perma.cc/R445-X79S], likely further complicate the calculus for OpenAI's superstakeholders, who may need to ask themselves what kind of deal they would accept from the Altman-led leadership in order to keep OpenAI out of Musk's hands — if, of course, they deem such a result desirable.

[184] *Cf.* Tallarita, *supra* note 20. ("When it comes to catastrophic risks, our legal system typically gives up on ordinary legal controls . . . and focuses on extraordinary legal controls, of the kind used to regulate nuclear proliferation or biohazard. The pursuit of AI safety warrants this kind of extraordinary effort . . . . While good corporate governance can help in the transitional phase, the government should quickly recognize its inevitable role in AI safety . . . .").