

Techstrong.ai

A Futurum Media Brand

SPECIAL REPORT / OCTOBER 3, 2026

Building AI, Fearing AI

The movements, money and personal rivalries behind AI's power networks



BY ALAN SHIMEL

Editor-in-Chief and CSO, Futurum Media

WHO CAN TELL THEM NO, AND MAKE IT STICK?

IN THIS ISSUE

The people behind the systems

AI's power is institutional as well as technical. This report follows the documented relationships that shape funding, careers, evaluation and governance.

- 03** **The builders' dilemma**
- 04** **Former colleagues and competing futures**
- 06** **The movements behind the leaders**
- 09** **Building AI to protect us from AI**
- 10** **Follow the money**
- 14** **How the network develops and places people**
- 16** **Who evaluates the evaluators**
- 17** **Power behind the scenes**
- 20** **When the label becomes a liability**
- 21** **Thiel and Palantir complicate the camps**
- 22** **Who can tell them no**

THE EVIDENCE, WHERE YOU NEED IT

Definitions and dated connections appear within the chapters.
23 Methodology · 24 Notes and sources

THE QUESTION

The builders' dilemma

If you believe the technology you are building could threaten humanity, why keep building it?

That is the question I want answered when an AI executive warns about the danger of increasingly powerful artificial intelligence. The warning may be sincere. The danger may be real. Neither explains, by itself, why the person delivering it is also competing to build the next system.

In May 2023, Sam Altman and Dario Amodei signed a statement identifying AI extinction risk as a global priority comparable to pandemics and nuclear war. Elon Musk backed a separate call for a six-month pause on training systems more powerful than GPT-4. He also developed an AI competitor of his own. These concerns were coming from people involved in advancing the technology, with resources and decisions under their control.^{1,2,3}

Amodei supplies a useful answer to the apparent contradiction. In his October 2024 essay, "Machines of Loving Grace," he describes enormous potential gains in health, economic development and other areas. He regards risks as obstacles to a fundamentally positive future, not reasons to abandon that future.⁴

You can understand that position without accepting the judgment behind it. A developer can believe a technology is both valuable and dangerous. The harder questions concern how much danger is acceptable, who gets to decide and what happens when the developer is wrong.

Yann LeCun represents a different approach. He signed a 2023 statement making the case that openness, transparency and wider access are essential to AI safety. It warned that the wrong regulation could concentrate power and undermine competition. The dispute therefore reaches beyond how dangerous AI might become. It concerns who should have access to it and who should govern it.⁵

The people arguing these positions have histories with one another. Musk and Altman helped establish OpenAI. Amodei worked there before leaving to build Anthropic. Former colleagues became competitors. Public arguments about humanity's future now arrive through companies competing for customers, capital and influence.

But the founders are only the most visible part of the story. Behind them sit donors, career organizations, research institutes, evaluators and governance bodies. Some share funders. Some share personnel. Some possess formal appointment rights. Their activities help determine which questions receive attention, which careers get supported and whose assessment becomes authoritative.

These groups deserve examination as institutions exercising power, not simply as communities exchanging ideas. The public record supports a network of consequential relationships. It does not establish a single hidden command center running AI. We can investigate the former without pretending we have proved the latter.

1. <https://safe.ai/work/statement-on-ai-extinction-risk>

2. <https://futureoflife.org/newsletter/future-of-life-institute-newsletter-pause-giant-ai-experiments/>

3. <https://www.reuters.com/legal/elon-musk-sues-openai-ceo-sam-altman-breach-contract-2024-03-01/>

4. <https://darioamodei.com/essay/machines-of-loving-grace>

5. <https://open.mozilla.org/ai-safety/>

THE FOUNDERS

Former colleagues and competing futures

The Musk–Altman relationship is a good place to begin because its breakdown exposes a question running through the entire report: Can someone claim to be protecting humanity while also seeking control over the institution doing the protecting?

Musk's 2024 lawsuit accused OpenAI and Altman of abandoning the organization's original public-benefit mission in favor of commercial interests. That was an allegation, not an established finding. OpenAI countered with its own account and released communications that it said showed Musk had supported a for-profit structure while seeking control.^{3,6}

In OpenAI's telling, disagreements included majority equity, board control and a proposed combination with Tesla. The company argued that allowing one person absolute control would have conflicted with its mission. Those are the claims of a party to the dispute, and they should be read accordingly.⁶

The legal history has since moved on. In May 2026, a jury in Musk's later federal case found that he had sued too late; Reuters reported that Musk said he would appeal. That verdict is not a finding that every disputed account of OpenAI's founding history was true or false.³⁷

What the public exchange reveals is that "benefit humanity" does not resolve an ownership dispute. A mission can leave open who chooses the strategy, appoints the board, supplies the capital and decides when the mission has been fulfilled. Former partners can agree that AI matters enormously while disagreeing about whose institution should lead.

The Altman–Amodei split produced a different competitor. Anthropic's founding team included former OpenAI researchers. Its 2021 financing announcement positioned reliability and safety as central to its work.^{7,8}

"'Benefit humanity' does not resolve an ownership dispute."

ALAN SHIMEL

3. <https://www.reuters.com/legal/elon-musk-sues-openai-ceo-sam-altman-breach-contract-2024-03-01/>

6. <https://openai.com/index/openai-elon-musk/>

7. <https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>

8. <https://www.theatlantic.com/technology/2026/09/openai-v-anthropic-inside-biggest-rivalry-tech/688819/>

37. <https://www.reuters.com/legal/government/elon-musk-loses-lawsuit-against-openai-2026-05-18/>

THE INSTITUTIONAL RECORD

Trust becomes institutional

Recent reporting adds a more personal account. The Atlantic's September 2026 reconstruction describes distrust and disagreements over commercialization behind the founders' departure. In his June 17 Bloomberg interview, Amodei discussed trust in explaining the split and cited his relationship with Demis Hassabis as evidence that cooperation among competitors was possible. These are attributed accounts, not independent findings about the participants' character or private thinking.^{8,9}

Trust has practical consequences in a field asking competitors to share information about systems they may not fully understand. A company can support safety cooperation in principle and still question whether another company will disclose inconvenient findings. It can favor independent review while disputing who qualifies as independent.

That puts personal history close to institutional design. Who gets invited into a laboratory? Whose assurances are accepted? Who is regarded as reckless, competent or honest? Answers can affect cooperation even when nobody writes them into a charter.

LeCun's position adds another source of friction. The openness statement he signed rejects the proposition that tight proprietary control is the sole route to protection.⁵ That argument challenges a model in which a small number of companies develop advanced systems and negotiate oversight largely among themselves.

None of these relationships should be reduced to a soap opera. Nor should they be scrubbed out of the story in favor of an abstract debate about alignment. Institutions are built by people who choose collaborators, leave employers, back founders and decide whom to trust. Those decisions can outlast the friendship or professional partnership that produced them.

For readers buying AI products, the consequence is straightforward. The laboratory supplying your system has a technical approach, a commercial strategy and a governance history. Its competitors may share some scientific assumptions while disagreeing sharply about leadership and oversight. "The AI industry believes" is often too broad a description to be useful.



AI-generated conceptual illustration

5. <https://open.mozilla.org/ai-safety/>

8. <https://www.theatlantic.com/technology/2026/09/openai-v-anthropic-inside-biggest-rivalry-tech/688819/>

9. <https://www.youtube.com/watch?v=x2VHFgyawPE>

THE MOVEMENTS

The movements behind the leaders

The argument over AI is not just a contest among companies or a quarrel among founders. It is also a dispute over what technological progress is for, whose welfare should count and how much risk society should accept. Before following the money and institutions behind these positions, we need to separate the ideas that are too often bundled together.

Effective altruism, usually shortened to EA, begins with a reasonable proposition: Use evidence and reasoning to identify effective ways to help others. Its work includes global health, animal welfare and pandemic prevention as well as AI. Describing it only as an AI-doom movement would erase much of what its participants actually do.¹⁰

The movement's broader ambition is to compare causes and interventions rather than accept that every worthwhile project deserves the same priority. That invites difficult choices. Money spent addressing one problem cannot be spent on another. A career devoted to one field is a career not devoted to something else.

Longtermism introduces another consideration: the welfare of people who may live far into the future. It is a moral view, not a synonym for EA. In AI, it can make avoiding a catastrophe especially important because the consequences could extend beyond the people alive today.¹¹

There is an enormous difference between estimating the effectiveness of a present-day health intervention and assigning probabilities to a civilization-ending technology scenario. Both may involve evidence and judgment. They do not have the same opportunities for measurement or verification. A spreadsheet does not eliminate the assumptions used to populate it.

That is where scrutiny belongs. Which assumptions dominate the calculation? How sensitive is a recommendation to a different forecast? What evidence would change the conclusion? People affected by the resulting priorities deserve intelligible answers, not just assurances that the reasoning is rigorous.

READING THE MOVEMENTS

Effective altruism	A movement and practical project seeking effective ways to help others through evidence and reasoning. ¹⁰ Includes non-AI causes; not a single governing body.
Longtermism	A moral view giving substantial weight to the long-run future and future generations. ¹¹ Overlaps with EA but is not identical to it.

10. <https://www.effectivealtruism.org/articles/introduction-to-effective-altruism>

11. <https://www.effectivealtruism.org/articles/longtermism>

THE INSTITUTIONAL RECORD

Communities, fields and futures

Rationalism is another overlapping community, associated with platforms such as LessWrong and an interest in improving reasoning and decision-making. It is not the organizational headquarters of all EA activity. Participation in a discussion community does not establish agreement with everything its prominent writers believe.¹²

AI safety and alignment, meanwhile, are fields of work. Researchers can study harmful behavior, misuse or loss of control without identifying with EA or longtermism. Treating every safety engineer as a movement adherent would be as careless as assigning a political identity to everyone working in cybersecurity.

On the other side of many public arguments are effective accelerationism, or e/acc, and techno-optimism. Early e/acc writings favor accelerated technological development. Marc Andreessen's "Techno-Optimist Manifesto" offers a broader argument for technological progress, markets and growth, while attacking what it considers obstacles to advancement. These are related currents, not interchangeable organizations.^{13,14}

Transhumanism supplies a further strand: using technology to extend or enhance human capabilities. It helps explain some ambitions surrounding longevity and human transformation, but it should not be attached automatically to every optimistic AI forecast.¹⁵

READING THE MOVEMENTS

Rationalism	A community concerned with improving reasoning and decision-making, associated with LessWrong. ¹² Forum participation is not an allegiance test.
AI safety and alignment	Technical and policy work addressing AI harms and intended behavior. Work in the field does not establish movement identity.
Effective accelerationism	A current favoring accelerated technological development. ¹³ Not synonymous with every pro-innovation position.
Techno-optimism	Broad confidence in the benefits of technological progress; Andreessen's manifesto is one formulation. ¹⁴ Not one organization or a uniform policy program.
Transhumanism	Ideas about technological enhancement or extension of human capabilities. ¹⁵ Historical overlap does not establish an individual's affiliation.

12. <https://www.lesswrong.com/about>

13. <https://beff.substack.com/p/notes-on-eacc-principles-and-tenets>

14. <https://a16z.com/the-techno-optimist-manifesto/>

15. <https://nickbostrom.com/ethics/values>

THE INSTITUTIONAL RECORD

Ideas become institutions

The labels matter less than the organizations built around them. A philosophy acquires another kind of importance when it helps guide a funding portfolio, a career recommendation or the selection of a corporate director. At that point, we are examining decisions with consequences outside the community.

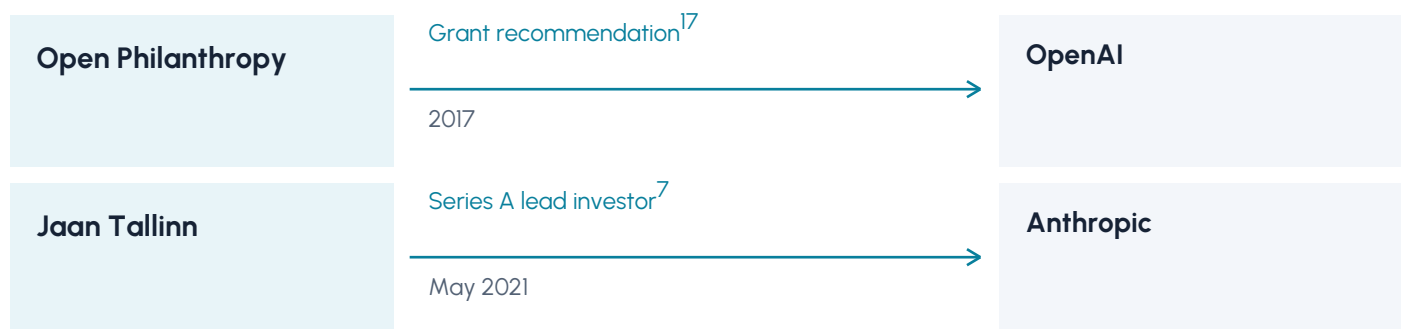
There is no single membership test that settles this investigation. A person may reject the EA label and agree with particular longtermist arguments. Another may identify with EA while concentrating on present-day poverty. Similar language is a reason to examine intellectual influence, not proof of concealed allegiance.

Why, then, do so many insiders sound alike? There are several explanations worth separating. Researchers studying similar systems may independently reach similar conclusions. Colleagues may learn a framework from one another. Employers may select people already interested in the questions they prioritize. A successful funding program may expand the number of researchers working within a particular approach.

Those mechanisms can operate together. None requires a script distributed by a central authority. To distinguish them, we would need more than a list of overlapping quotations: dated research, accounts of how views developed, selection criteria and evidence about competing approaches that did or did not receive support.

The possibility of independent agreement is important. So is the possibility that a well-funded intellectual community makes its conclusions more visible than alternatives. Counting public warnings cannot tell us which explanation dominates. Looking at the institutions producing and amplifying them gets us closer.

THE COMPANY CONNECTIONS



Dated relationships, not ideological affiliations or a common chain of command.

7. <https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>

17. <https://www.openphilanthropy.org/research/our-progress-in-2017-and-plans-for-2018/>

THE INCENTIVES

Building AI to protect us from AI

The most persuasive explanation for the builders' dilemma is also the one that requires the toughest accountability: The benefits could be enormous, the risks require intervention, and the people developing the systems believe they can help make the outcome better.

Amodei's writing makes clear that optimism and concern can occupy the same worldview.⁴ It is possible to believe that stopping development would sacrifice valuable advances while continuing without safeguards would be irresponsible. The practical dispute is about the safeguards and whether they keep pace with capabilities.

In his September 2026 pacing essay, Amodei argues that slower development could give safety work more time. He also considers commercial advantage and the United States' lead over China. He proposes embedded external reviewers with substantial access and rights to publish findings, subject to specified redactions. Those are proposals and commitments whose implementation must be assessed separately.¹⁶

That position exposes the difficulty of voluntary restraint. If every company says it must remain ahead of less responsible competitors, the argument can justify continued expansion indefinitely. Each firm can regard itself as the necessary exception. A safety commitment needs a stopping condition that survives the arrival of a promising business opportunity.

Developers also have a legitimate need to test systems and investigate their behavior. But "we need to build it to understand it" cannot be an unlimited authorization. Testing arrangements, access permissions and deployment boundaries matter. Research can justify an experiment without justifying broad release.

The same distinction applies to geopolitical competition. National-security concerns may be genuine. They still require evidence connecting a proposed restriction or acceleration to a defensible outcome. Invoking another country does not settle whether a particular company should receive more discretion.

Warnings can also enhance the authority of the person delivering them. A developer who identifies a danger may become the expert governments consult about its solution. That can happen without a cynical motive. It is nevertheless a conflict worth recognizing when the proposed solution affects competitors, customers and public policy.

For enterprise buyers, promises become useful when they produce verifiable obligations. Does a safety finding delay a release? Can an outside reviewer publish a damaging result? Who can authorize deployment over an objection? Can a customer obtain the evidence relevant to its own use case?

The love-hate relationship with AI becomes consequential at those decision points. A compelling vision of the future is not a substitute for a process that can reject a risky product in the present.

4. <https://darioamodei.com/essay/machines-of-loving-grace>

16. <https://darioamodei.com/post/we-must-pace-the-frontier>

THE MONEY

Follow the money

The financial record shows institution-building, not merely funding for isolated papers. To follow it accurately, we have to distinguish a donor from an adviser, a grant from an investment and an announced commitment from cash received.

An Open Philanthropy retrospective reports a \$30 million OpenAI grant recommendation within its 2017 advanced-AI-risk work. OpenAI's February 2018 supporters announcement listed Holden Karnofsky, then an Open Philanthropy leader, on its board and identified Jaan Tallinn among its donors.^{17,18}

Those records put philanthropic support and corporate governance in the same historical picture. They do not tell us that a funder directed a particular model release. They do show why an account of early AI institutions that includes only venture capital and founders is incomplete.

Georgetown announced a \$55 million Open Philanthropy grant when CSET launched in 2019. CSET subsequently announced an additional \$42 million in 2021.^{19,20} This was funding for an institution positioned to provide policy research, not an equity purchase in an AI laboratory.

Anthropic's financing belongs in a different category. Its May 2021 Series A totaled \$124 million, led by Tallinn and including Dustin Moskowitz among the participants. Its April 2022 Series B totaled \$580 million, led by Sam Bankman-Fried, with participants including Caroline Ellison and Nishad Singh. These are round totals, not individual contribution amounts, and they are investments rather than charitable grants.^{7,21}

Bankman-Fried's presence brings the FTX story into the report, but sloppy accounting would turn it into a misleading one. An investment in a company and a foundation's grant to a charity create different obligations. They cannot be combined under a single label and used to imply identical influence.

SELECTED GRANTS / GRANT RECOMMENDATIONS



USD millions. Common zero baseline; scale 0–60. Not cash-receipt totals.^{17,19,20}

7. <https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>

17. <https://www.openphilanthropy.org/research/our-progress-in-2017-and-plans-for-2018/>

18. <https://openai.com/index/openai-supporters/>

19. <https://www.georgetown.edu/news/largest-u-s-center-on-artificial-intelligence-policy-comes-to-georgetown/>

20. <https://cset.georgetown.edu/article/new-grant-agreement-boosts-cset-funding-to-more-than-100-million-over-five-years/>

21. <https://www.anthropic.com/news/anthropic-raises-series-b-to-build-safe-reliable-ai>

THE MONEY / THE INSTITUTIONAL CONNECTIONS

Where the funding leads

OpenAI, Anthropic and policy research enter the same institutional picture through different kinds of support. A grant, directorship and equity investment are not interchangeable.



Dated relationships, not ideological affiliations or a common chain of command.

7. <https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>

17. <https://www.openphilanthropy.org/research/our-progress-in-2017-and-plans-for-2018/>

18. <https://openai.com/index/openai-supporters/>

19. <https://www.georgetown.edu/news/largest-u-s-center-on-artificial-intelligence-policy-comes-to-georgetown/>

20. <https://cset.georgetown.edu/article/new-grant-agreement-boosts-cset-funding-to-more-than-100-million-over-five-years/>

THE INSTITUTIONAL RECORD

How funding shapes a field

In December 2023, Effective Ventures reported returning \$26,786,503 to the FTX bankruptcy estate, equal to its stated 2022 receipts from FTX and the FTX Foundation. It also said an outside investigation found no awareness of criminal fraud among its trustees or staff. That investigation was not published with the announcement. The organization's response belongs beside the financial disclosure.²²

More recent allocation machinery is visible through the Survival and Flourishing Fund. Its 2026 main-round announcement identifies Tallinn and Moskovitz as funders and explains a process in which funders can adjust intended allocations.²³ We do not need to speculate that donor preferences exist. The allocation process acknowledges them.

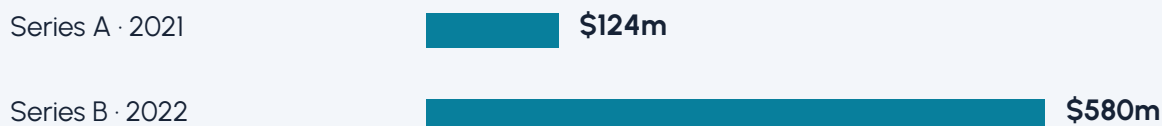
The more difficult task is determining how those preferences affect a recipient's subsequent work. A donor can choose a field without choosing every result. A recipient can retain editorial or scientific independence while remaining dependent on renewal. The conditions in the agreement, the availability of alternative funding and the history of disagreement all matter.

This is where the network can exercise power without issuing instructions. Funding determines which proposed institution gets staff, which program survives and which research question receives years of attention. A donor does not have to rewrite a paper to influence the supply of papers available to policymakers.

Private philanthropy can also support work that commercial markets neglect. That is a substantial benefit, particularly for research whose value depends on preventing harm rather than selling a product. The accountability question is whether the public can inspect the concentration of support and understand the boundaries of donor authority.

For that reason, the useful funding table is dated and specific. It should not offer an impressive-looking total assembled from unlike categories. It should identify the origin, recipient, purpose, terms where available and what the underlying disclosure actually establishes.

ANTHROPIC / HISTORICAL EQUITY ROUND TOTALS



USD millions; scale 0–600. Total rounds, not individual contributions.^{7,21}

7. <https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>

21. <https://www.anthropic.com/news/anthropic-raises-series-b-to-build-safe-reliable-ai>

22. <https://forum.effectivealtruism.org/posts/HjsfHwqasyQMWRzZN/ev-updates-ftx-settlement-and-the-future-of-ev>

23. <https://survivalandflourishing.fund/2026/recommendations>

THE MONEY / READING THE RECORDS

Follow the category

The funding discussed in this report, by category and date. These records are not an ecosystem total. Do not sum grants, investments and returned funds.

\$30 million

2017

Open Philanthropy → OpenAI¹⁷

Grant recommendation reported in retrospective; underlying agreement and payment schedule not recovered.

\$55 million

2019

Open Philanthropy → Georgetown CSET¹⁹

Founding grant announced by university.

\$42 million

2021

Open Philanthropy → Georgetown CSET²⁰

Enhanced grant agreement; not an additional \$100 million award.

\$124 million

May 2021

Tallinn-led investor group → Anthropic⁷

Total Series A; individual allocations not disclosed.

\$580 million

April 2022

Bankman-Fried-led investor group → Anthropic²¹

Total Series B; equity financing, not a foundation grant.

\$26,786,503

December 2023 disclosure

Effective Ventures → FTX bankruptcy estate²²

Returned funds, equaling EV's stated 2022 receipts.

\$250,000

2013 review period

Peter Thiel → MIRI³⁴

Donation reported April 2014; March-to-March reporting period, not a calendar year.

7. <https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>

17. <https://www.openphilanthropy.org/research/our-progress-in-2017-and-plans-for-2018/>

19. <https://www.georgetown.edu/news/largest-u-s-center-on-artificial-intelligence-policy-comes-to-georgetown/>

20. <https://cset.georgetown.edu/article/new-grant-agreement-boosts-cset-funding-to-more-than-100-million-over-five-years/>

21. <https://www.anthropic.com/news/anthropic-raises-series-b-to-build-safe-reliable-ai>

22. <https://forum.effectivealtruism.org/posts/HjsfHwqasyQMWRzZN/ev-updates-ftx-settlement-and-the-future-of-ev>

34. <https://intelligence.org/2014/04/02/2013-in-review-fundraising/>

THE TALENT

How the network develops and places people

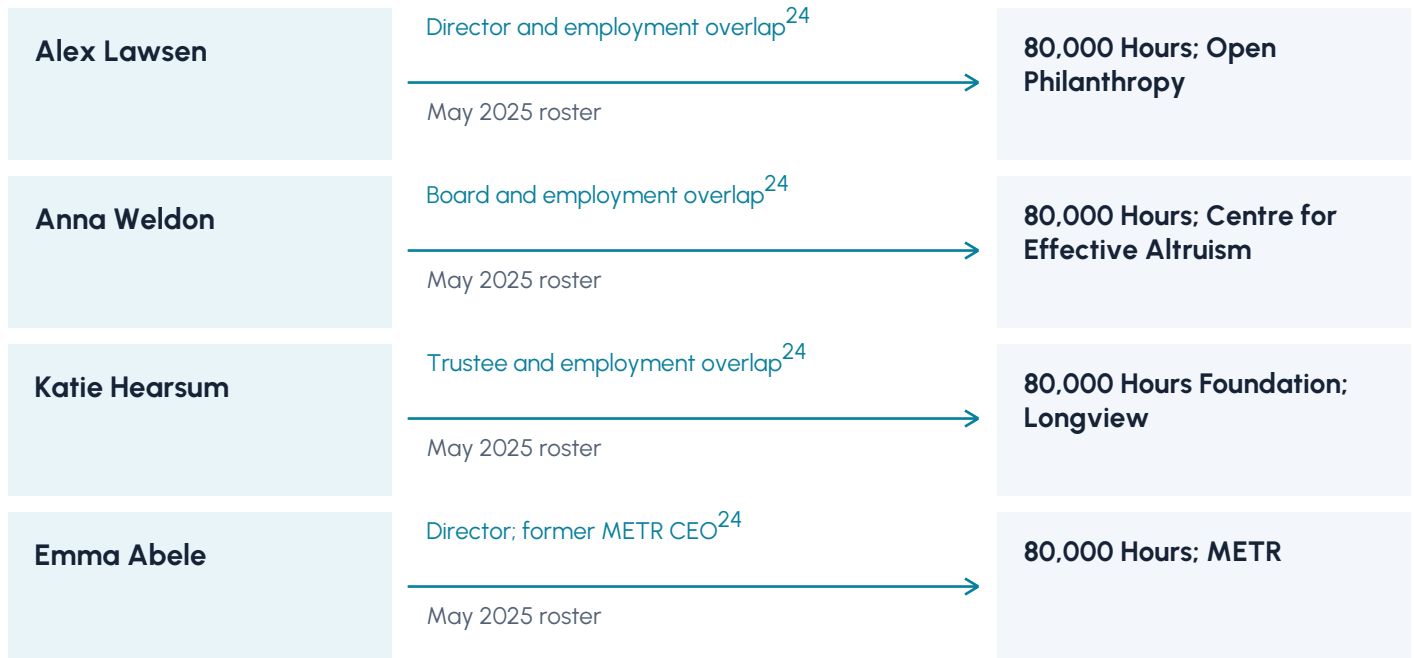
Money alone does not staff a research field. Institutions need people, and career programs can influence who enters the field, what they study and which organizations they encounter first.

80,000 Hours provides a documented example of continuing relationships across legally separate organizations. Its May 2025 spinout announcement described new governance arrangements and directors with roles at Open Philanthropy, the Centre for Effective Altruism, Longview and the UK AI Safety Institute. It also included a former METR chief executive.²⁴

Those were dated appointments, not a claim about today's roster. They reveal a professional network connecting career guidance, philanthropy, evaluation and public institutions. Legal separation does not erase shared experience. Shared experience does not establish a common chain of command.

Constellation's Astra fellowship makes the talent-development purpose explicit. Its program page describes funded research and reports that more than 80% of its first cohort moved into full-time AI safety roles. That is the program's own claim, not an independently audited placement rate.²⁵

THE CONNECTIONS IN THIS CHAPTER



Dated relationships, not ideological affiliations or a common chain of command.

24. <https://80000hours.org/2025/05/80000-hours-completes-spin-out-from-effective-ventures/>

25. <https://constellation.org/programs/astra>

THE INSTITUTIONAL RECORD

Who gets a route into the field

A fellowship can lower the cost of entering a difficult field, provide mentorship and help employers find capable researchers. It can also become an important point of access when applicants have few other routes to equivalent training. The selection process then deserves attention alongside the advertised curriculum.

The questions are practical. Are applicants assessed for skills, research interests or agreement with a worldview? Can a fellow conclude that a dominant risk theory is mistaken? Are grants available for approaches that challenge the program's premises? What happens when the mentor and the eventual employer come from the same small professional circle?

The records examined for this report do not establish systematic ideological hiring requirements or punishment for dissent. They do establish organizations intentionally developing talent and relationships spanning several kinds of institution.

That distinction allows a more useful inquiry than calling the community a secret society. Many of the programs and appointments are public. Their importance may be overlooked because individual disclosures are scattered across websites rather than presented together.

A healthy field needs expertise and continuity. It also needs routes for people without the right introduction. Competitive funding, transparent selection standards and support for competing research approaches would help demonstrate that the network develops knowledge rather than simply reproducing its own assumptions.

The relevant measure is not how many participants share a label. It is whether the field's opportunities remain open to qualified people who disagree with the people providing them.

>80%

Astra's first cohort moved into full-time AI safety roles, according to the program. Self-reported; not independently audited.²⁵

THE ACCOUNTABILITY TEST: Are selection standards transparent, and can fellows challenge the program's premises?

25. <https://constellation.org/programs/astra>

THE EVALUATORS

Who evaluates the evaluators

An evaluator gains importance when its work helps determine whether a model is considered safe enough for a consequential use. At that point, its independence is not a matter of branding. It affects decisions made by companies, governments and customers.

METR says it has not accepted AI-company funding and cannot accept donations made by or at the direction of frontier AI-company employees. It also discloses receiving substantial free model tokens.²⁶ Financial separation therefore needs to be considered alongside access to the systems being studied.

Access is necessary for serious evaluation. It can also create dependence. An evaluator needs to know what version it received, which capabilities were available, what restrictions applied and whether the environment resembled deployment. A company withholding relevant access can weaken an assessment without attempting to influence its written conclusion.

METR's August 2026 conflict policy recognizes personal connections as well as financial interests. It covers employment, holdings, family and romantic relationships, cohabitation and close friendships, with disclosure and mitigation requirements.²⁷ That is a useful acknowledgment that professional independence can be affected by relationships outside a funding agreement.

The existence of a policy does not establish how consistently it is applied. Nor does the existence of a friendship invalidate a finding. A defensible review needs enough disclosure for outsiders to understand the risk of bias and enough documentation to assess the response.

There is also intellectual independence. Two organizations can maintain separate finances and still approach a problem using the same assumptions. That may reflect strong evidence. It may also leave some failure modes underexamined. Competing methods and independent replication are ways to distinguish those possibilities.

Embedded reviewers could improve the situation by seeing more of what a company actually does. But greater proximity creates its own design questions. The reviewer needs protected publication rights, a clear way to report denied access and recusal procedures that work even when expertise is scarce.

A public assessment should also define its scope. A finding about one capability is not a certificate that a product is safe for every application. Enterprise deployments have different permissions, data and consequences. A laboratory test and a customer's operational risk assessment serve different purposes.

The network should welcome examination of its reviewers for the same reason it advocates examination of developers. If an assessment is important enough to influence deployment or regulation, its institutional foundations are important enough to disclose.

26. <https://metr.org/about>

27. <https://metr.org/coi-policy.pdf>

THE AUTHORITY

Power behind the scenes

The strongest evidence of institutional power is a decision someone has the authority to make. A board appointment is more concrete than a shared philosophy. A funded position is more concrete than a social-media exchange. A published policy proposal shows advocacy; it does not, by itself, prove control over the resulting law.

Anthropic supplied a clear governance example in April 2026. It announced that its Long-Term Benefit Trust had appointed Novartis chief executive Vas Narasimhan to the board and that trust-appointed directors then formed a majority. The company described the trust as independent, with members holding no financial stake in Anthropic.²⁸

That announcement documents exercised appointment power. It does not establish that the appointed directors share one ideology, or that the same arrangement will survive every subsequent corporate change. But it shows an institution acting through formal rights rather than merely offering advice.

The public-benefit purpose deserves consideration. A governance body insulated from immediate financial returns may provide a counterweight to investor pressure. Yet a mandate to represent humanity raises questions about selection and accountability of its own. Who appoints the trustees? Who can remove them? What information can affected people obtain? How are disagreements resolved?

THE CONNECTIONS IN THIS CHAPTER



Dated relationships, not ideological affiliations or a common chain of command.

28. <https://www.anthropic.com/news/narasimhan-board>

THE INSTITUTIONAL RECORD

Policy, boards and journalism

Policy research is another route to influence. CSET's 2021 announcement described former staff working at the National Security Council, the Office of Science and Technology Policy and Congress. It also described fellowships and research cited by public institutions. CSET said its agenda was shaped by experts and its methods were evidence-based.²⁰

The documented result is institutional capacity and access. It would take a separate reconstruction of a specific policy decision to establish that a donor directed its outcome. A staffing pipeline into government should be visible without being portrayed as proof that everyone who passes through it follows private instructions.

Helen Toner's biography and historical OpenAI appointment illustrate movement among grantmaking, policy research and corporate governance. Her background included Open Philanthropy and CSET; OpenAI announced her board appointment in September 2021. That is a historical connection, not a continuing board seat.^{29,30}

Journalism belongs in the picture as well. Tarbell's published policies say that most of its funding originated with Coefficient Giving as of 2025. They also assign editorial control to supported journalists and newsrooms, require disclosures and prohibit donor approval of journalism or personnel.³¹

THE CONNECTIONS IN THIS CHAPTER



Dated relationships, not ideological affiliations or a common chain of command.

20. <https://cset.georgetown.edu/article/new-grant-agreement-boosts-cset-funding-to-more-than-100-million-over-five-years/>

29. <https://cset.georgetown.edu/staff/helen-toner/>

30. <https://openai.com/index/helen-toner-joins/>

31. <https://www.tarbellcenter.org/ethics-and-standards>

THE INSTITUTIONAL RECORD

Influence is not total control

Supporting reporting on a complex industry can serve the public. Concentrated funding still merits examination. A reasonable test is whether supported journalism can scrutinize the funder and its associated institutions, with affiliations disclosed and without intervention. A grant alone does not establish purchased coverage.

Taken together, these cases reveal several forms of power operating behind the best-known executives. Some institutions choose where money goes. Others help people enter influential positions. Others possess appointment rights or supply the analysis decision-makers consult.

The network need not be centrally directed to matter. Its organizations can pursue separate strategies while collectively increasing the prominence of certain questions and approaches. That is an analytical interpretation of the documented relationships, not evidence of a coordinated plan.

For readers, the distinction is worth preserving. We should be able to identify substantial influence before it becomes total control. Otherwise, the only two available descriptions are an innocent collection of friends or an all-powerful conspiracy. Neither is an adequate account of institutions.

Consider a hypothetical policy recommendation supported by several organizations. Agreement looks more persuasive if the organizations reached it independently, using different methods and funding sources. It looks different if their work draws on the same underlying study, their advisers overlap and one donor supplies most of their budgets. The recommendation could still be correct. The number of organizational logos would simply overstate the number of independent judgments.

That is why tracing relationships must go beyond naming people who know one another. It should establish what was decided, which evidence was used and what dependencies were relevant. A regulator should be able to distinguish five separate assessments from five repetitions of one assessment. A customer should be able to do the same with safety claims.

This is also a fairer standard for the institutions themselves. A specific account of authority and dependence can be answered with records. A vague accusation about an invisible cabal cannot. Public scrutiny works better when the claim is precise enough to verify, refute or correct.

WHAT THE RECORD CAN ESTABLISH

**A grant. An appointment.
A contract. A disclosed role.**

What it cannot establish on its own: donor direction of a particular law, assessment or article.

THE IDENTITY

When the label becomes a liability

The question of why people distance themselves from EA has evidence behind it, but not a universal answer.

Rethink Priorities' December 2022 community survey, conducted with the Centre for Effective Altruism, found that 38% of respondents to the FTX questions were less likely to associate publicly with EA after the crisis. Thirty-one percent reported substantial loss of trust in public figures or leadership, excluding FTX and Alameda employees. The survey was self-selected and skewed toward engaged participants.³²

That is historical evidence of reluctance within a community. It is not a current count of AI workers concealing an affiliation. A separate 2024 U.S. public survey estimated awareness of EA at 1–2% under permissive criteria and below 1% under stringent criteria, with mostly positive views among those aware. Insider reputational damage should not be confused with universal public stigma.³³

For any individual, distancing could reflect disagreement with the movement's philosophy, disappointment in its institutions, rejection of the label or a preference for a broader corporate identity. Those possibilities require different evidence. Continued contact with former colleagues does not establish that a denial is dishonest.

Nor does rejecting the name make an institutional relationship disappear. An executive's self-description answers one question. A grant agreement, board appointment or advisory role answers another. The report should examine both without treating either as a substitute for the other.

EA has no single formal membership roll that resolves every case. "Officially part of it" is therefore a less useful test than asking what a person says, what they do and which documented responsibilities they hold.

The public does not need a declaration of philosophical identity from every engineer. It does need relevant disclosures from people making consequential decisions. A movement label is a poor substitute for transparent authority, funding and conflicts.

AFTER FTX / DECEMBER 2022 COMMUNITY RESPONSES

Less likely to associate publicly with EA

38%

Substantial loss of trust in EA leaders

31%

n=1,312 FTX-question respondents; self-selected, skewed toward engaged EAs. Bars: 0–100%. Trust excludes FTX/Alameda employees.³²

32. <https://rethinkpriorities.org/research-area/ftx-community-response-survey-results/>

33. <https://rethinkpriorities.org/research-area/pulse/>

ACROSS THE CAMPS

Thiel and Palantir complicate the camps

Peter Thiel does not fit neatly into a story of one safety movement opposed by one accelerationist movement. MIRI's 2013 fundraising review reports a \$250,000 donation from him. The review uses a March-to-March reporting period rather than a calendar year.³⁴ Historical support for AI-risk research can coexist with a distinct political philosophy.

Thiel's 2009 Cato essay also contains an explicit argument about tensions between freedom and democracy.³⁵ It is a dated statement by Thiel, not a platform that should be assigned to every company he helped establish or every recipient he supported.

The comparison worth making concerns how wealthy patrons build influence: financing institutions, backing founders and helping some visions acquire organizational support. Those methods can appear across competing philosophies. Similar methods do not prove identical goals.

Palantir belongs here because there is a documented corporate relationship, not because Thiel's presence automatically establishes one. In November 2024, Anthropic, Palantir and AWS announced a partnership bringing Claude models to U.S. government intelligence and defense operations through Palantir's platform and AWS infrastructure.³⁶

That announcement establishes commercial cooperation. It does not establish Palantir as an EA organization, a corporate funder of the movement or an executor of Thiel's personal philosophy. Nor should a historical announcement be presented as a complete account of the relationship today.

The partnership nevertheless illustrates why rigid camp boundaries can mislead. Companies can cooperate commercially while their leaders or associated patrons differ philosophically. Contracts connect institutions in ways that public rhetoric does not always capture.

For this investigation, the rule is to name the connection precisely. A donation, appointment, former job and commercial contract belong on the map with different descriptions. Putting the same unqualified line between every pair would manufacture an appearance of control the evidence does not support.

Peter Thiel → MIRI

Donation³⁴

2013 review period

Anthropic → Palantir; AWS

Commercial partnership³⁶

November 2024

Different relationships, not evidence of a shared ideology.

34. <https://intelligence.org/2014/04/02/2013-in-review-fundraising/>

35. <https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian>

36. <https://www.businesswire.com/news/home/20241107699415/en/Anthropic-and-Palantir-Partner-to-Bring-Claude-AI-Models-to-AWS-for-U.S.-Government-Intelligence-and-Defense-Operations>

THE ACCOUNTABILITY

Who can tell them no



AI-generated conceptual illustration

The builders' dilemma is not resolved by discovering that some participants are sincere and others have business interests. Sincerity and financial interest can coexist. So can genuine fear, ambition and a conviction that one's own organization is best equipped to lead.

The records establish donors supporting institutions, career programs developing talent, people moving across organizational boundaries and a trust exercising board appointment rights. They support examination of a power network whose importance reaches beyond public arguments among founders. They do not establish a unified secret society directing the industry.

Accountability should follow the actual authority. Funding disclosures should identify concentration and restrictions. Governance documents should explain appointment and removal rights. Evaluators should disclose relevant conflicts and limits on access. Policymakers should examine the origins of the expertise they rely on. Institutions claiming openness to criticism should provide workable routes for dissent.

Companies should also make safety commitments testable. If a reviewer can publish only with permission, or a deployment objection can always be overridden without explanation, the public has received reassurance rather than oversight.

We should not leave future generations out of the discussion. We should not leave people alive today outside the decision-making process either. The consequences reach customers, workers and communities that never agreed to adopt a founder's forecast or a donor's moral priorities.

The people building AI may love what it could accomplish and fear what it could become. Their institutions must be able to act on that fear when acting is inconvenient. And the networks behind them must be visible enough to challenge.

Who can tell them no, and make it stick?

HOW TO READ THIS REPORT

Evidence, scope and limits

This report draws on public organizational disclosures, announcements, published policies, survey reports and attributed news reporting reviewed through October 3, 2026. No interviews, requests for comment or independent financial audit were conducted. Historical disclosures establish relationships at the dates indicated. Organizational program outcomes are identified as self-reported. Editorial interpretations are distinguished from claims of formal authority or coordinated conduct. The selected records do not establish movement prevalence among AI workers, systematic ideological gatekeeping or centralized covert control. AI-generated images are conceptual illustrations, not photographs of actual events, people or facilities. Data graphics are constructed from the cited records, not generated by an image model.

Documented

A source identifies the amount, role, policy, appointment or commercial relationship.

Attributed

An organization's self-description, a participant's recollection or a reported allegation is not an independent finding.

Analytical

The report's argument concerns how institutions can exercise influence. It does not claim a centrally directed network.

Historical

Dates belong to the evidence. A past directorship or announced partnership must not be read as a current office or contract.

Techstrong.ai

A Futurum Media Brand

Written by Alan Shimel

Editor-in-Chief and CSO, Futurum Media

October 3, 2026 · Urbanist typography

Original AI-generated conceptual artwork

THE SOURCE RECORD

Notes and sources

Notes 1–11 · Source pages checked October 3, 2026.

- 01** Center for AI Safety, Statement on AI Extinction Risk, May 2023 statement and signatories.
<https://safe.ai/work/statement-on-ai-extinction-risk>
- 02** Future of Life Institute, Pause Giant AI Experiments newsletter, March 31, 2023. The institute names Musk as a signatory.
<https://futureoflife.org/newsletter/future-of-life-institute-newsletter-pause-giant-ai-experiments/>
- 03** Reuters, Elon Musk sues OpenAI for abandoning original mission for profit, March 1, 2024. Used for the historical allegations and contemporaneous competing venture, not the litigation's current status.
<https://www.reuters.com/legal/elon-musk-sues-openai-ceo-sam-altman-breach-contract-2024-03-01/>
- 04** Dario Amodei, Machines of Loving Grace, October 2024.
<https://darioamodei.com/essay/machines-of-loving-grace>
- 05** Mozilla and signatories, Joint Statement on AI Safety and Openness, October 31, 2023. LeCun appears among the signatories.
<https://open.mozilla.org/ai-safety/>
- 06** OpenAI, OpenAI and Elon Musk, March 5, 2024. A party's account of disputed history, not an independent finding.
<https://openai.com/index/openai-elon-musk/>
- 07** Anthropic, Series A announcement, May 28, 2021.
<https://www.anthropic.com/news/anthropic-raises-124-million-to-build-more-reliable-general-ai-systems>
- 08** Kevin Roose, The Atlantic, Inside the Biggest Feud in Artificial Intelligence, September 29, 2026. Full retrieved article checked; used for a brief attributed account, not independent findings.
<https://www.theatlantic.com/technology/2026/09/openai-v-anthropic-inside-biggest-rivalry-tech/688819/>
- 09** Bloomberg Originals, Inside the Mind of Anthropic CEO Dario Amodei, The Circuit, extended interview, June 17, 2026. Transcript checked at 05:49–07:43 and 08:41–10:35; remarks are Amodei's account.
<https://www.youtube.com/watch?v=x2VHFgyawPE>
- 10** Effective Altruism, What is effective altruism.
<https://www.effectivealtruism.org/articles/introduction-to-effective-altruism>
- 11** Effective Altruism, Longtermism introduction.
<https://www.effectivealtruism.org/articles/longtermism>

THE SOURCE RECORD

Notes and sources

Notes 12–21 · Source pages checked October 3, 2026.

- 12** LessWrong, About.
<https://www.lesswrong.com/about>
- 13** Beff Jezos and contributors, Notes on e/acc principles and tenets.
<https://beff.substack.com/p/notes-on-eacc-principles-and-tenets>
- 14** Marc Andreessen, The Techno-Optimist Manifesto, October 2023.
<https://a16z.com/the-techno-optimist-manifesto/>
- 15** Nick Bostrom, Transhumanist Values.
<https://nickbostrom.com/ethics/values>
- 16** Dario Amodei, We Must Pace the Frontier, September 2026. Proposal and intended review arrangements are not presented as fully implemented controls.
<https://darioamodei.com/post/we-must-pace-the-frontier>
- 17** Open Philanthropy, Our Progress in 2017 and Plans for 2018, March 20, 2018. The retrospective reports grant recommendations, not an audited payment schedule.
<https://www.openphilanthropy.org/research/our-progress-in-2017-and-plans-for-2018/>
- 18** OpenAI, OpenAI supporters, February 20, 2018.
<https://openai.com/index/openai-supporters/>
- 19** Georgetown University, CSET launch announcement, February 28, 2019.
<https://www.georgetown.edu/news/largest-u-s-center-on-artificial-intelligence-policy-comes-to-georgetown/>
- 20** CSET, New grant agreement, August 25, 2021.
<https://cset.georgetown.edu/article/new-grant-agreement-boosts-cset-funding-to-more-than-100-million-over-five-years/>
- 21** Anthropic, Series B announcement, April 29, 2022.
<https://www.anthropic.com/news/anthropic-raises-series-b-to-build-safe-reliable-ai>

THE SOURCE RECORD

Notes and sources

Notes 22–29 · Source pages checked October 3, 2026.

- 22** Zachary Robinson, Effective Ventures, EV updates: FTX settlement and the future of EV, December 13, 2023, published on the EA Forum. The investigation's conclusions are the organization's account; the full investigation was not published with the post.
<https://forum.effectivealtruism.org/posts/HjsfHwqasyQMWRzZN/ev-updates-ftx-settlement-and-the-future-of-ev>
- 23** Survival and Flourishing Fund, 2026 main-round recommendations announcement.
<https://survivalandflourishing.fund/2026/recommendations>
- 24** 80,000 Hours, Spinout announcement and boards, May 28, 2025.
<https://80000hours.org/2025/05/80000-hours-completes-spin-out-from-effective-ventures/>
- 25** Constellation Institute, Astra fellowship, live program page. Outcome claim is self-reported.
<https://constellation.org/programs/astra>
- 26** METR, About and funding disclosures, live page.
<https://metr.org/about>
- 27** METR, Conflict of interest policy, version 1.0, August 28, 2026.
<https://metr.org/coi-policy.pdf>
- 28** Anthropic, Long-Term Benefit Trust appoints Vas Narasimhan, April 14, 2026. Used as a dated governance event, not a claim about future or post-IPO control.
<https://www.anthropic.com/news/narasimhan-board>
- 29** CSET, Helen Toner biography.
<https://cset.georgetown.edu/staff/helen-toner/>

THE SOURCE RECORD

Notes and sources

Notes 30–37 · Source pages checked October 3, 2026.

- 30** OpenAI, Helen Toner board appointment, September 8, 2021.
<https://openai.com/index/helen-toner-joins/>
- 31** Tarbell Center for AI Journalism, Ethics and Standards, including funding statement referring to 2025.
<https://www.tarbellcenter.org/ethics-and-standards>
- 32** Rethink Priorities, FTX Community Response Survey Results, December 2022 fieldwork, March 15, 2023 publication; 1,312 respondents to the FTX questions.
<https://rethinkpriorities.org/research-area/ftx-community-response-survey-results/>
- 33** Rethink Priorities, Pulse: US attitudes and awareness regarding effective giving and philanthropic cause areas, November 26, 2024; 4,890 U.S. adults after screening. Non-probability opt-in survey, weighted/poststratified estimates.
<https://rethinkpriorities.org/research-area/pulse/>
- 34** MIRI, 2013 in Review: Fundraising, April 2, 2014. Figures use a March-to-March reporting period, are approximate, and differ from calendar-year filings.
<https://intelligence.org/2014/04/02/2013-in-review-fundraising/>
- 35** Peter Thiel, The Education of a Libertarian, April 13, 2009.
<https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian>
- 36** Anthropic and Palantir partnership announcement, distributed by Business Wire, November 7, 2024. Claude access through Palantir and AWS for U.S. government intelligence and defense; historical announcement only.
<https://www.businesswire.com/news/home/20241107699415/en/Anthropic-and-Palantir-Partner-to-Bring-Claude-AI-Models-to-AWS-for-U.S.-Government-Intelligence-and-Defense-Operations>
- 37** Reuters, OpenAI defeats Elon Musk's lawsuit, removes obstacle to IPO, May 18, 2026, updated May 19. Jury found the later federal action untimely; Musk said he would appeal. No claim here about later appellate developments.
<https://www.reuters.com/legal/government/elon-musk-loses-lawsuit-against-openai-2026-05-18/>