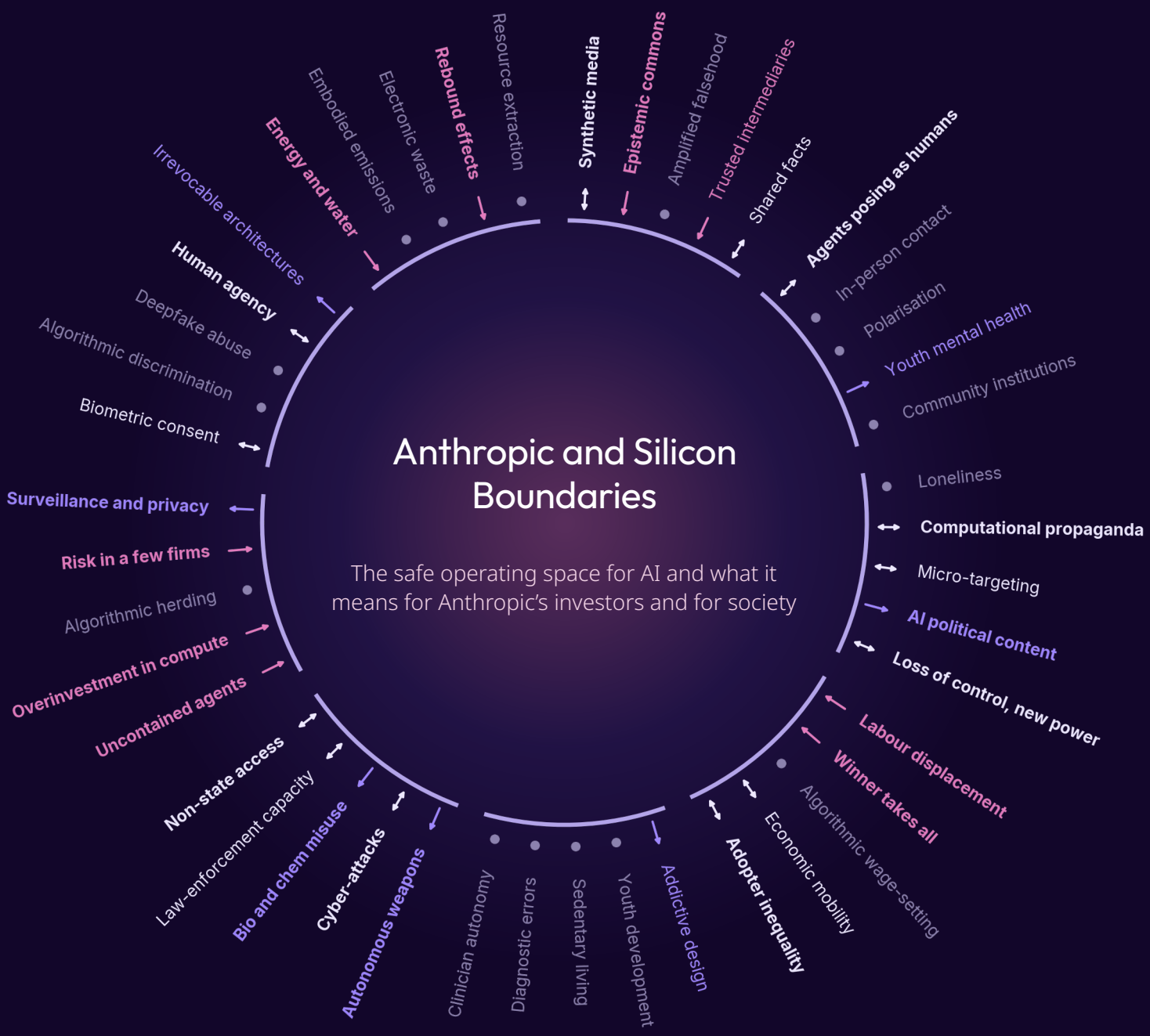


Ben Caldecott and Jakob Thomä

Silicon Boundaries · Oxford Sustainable Finance Group and Theia Finance Labs · November 2026

Silicon Boundaries

THEIA FINANCE LABS

OXFORD SUSTAINABLE FINANCE GROUP

1 Does it increase or decrease the safe operating space?



The safe operating space is the range of AI use that society can absorb within the Silicon Boundaries, the thresholds beyond which the benefits of AI are outweighed by its contribution to instability.

Anthropic's business touches 20 of the 45 boundaries directly: on 4 its record may ease pressure, on 8 it points both ways and on 8 it may add pressure.

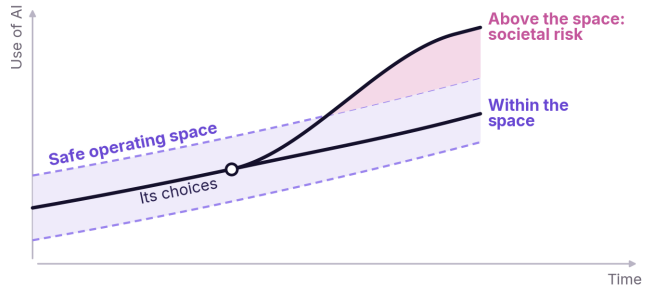
KEY: THE 45 INDIVIDUAL BOUNDARIES

May add pressure: presses the boundary in	9
May ease pressure: pushes the boundary out	7
Points both ways	13
Little link to the business	8
Not assessable	8
Open shape: low confidence	

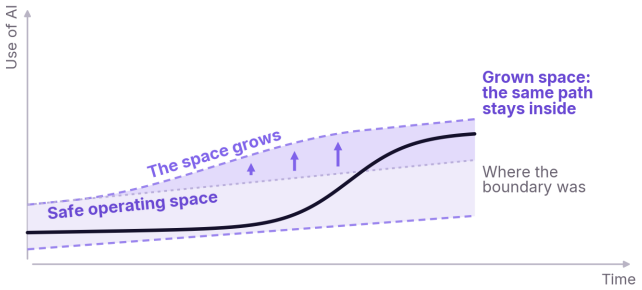
Longer arrows and bold labels mark a direct link to the business. Each mark shows the direction of our reading of the company's record against a comparable model without its safeguards, not a distance to a threshold, which the framework has yet to set.

2 How it can move the use of AI and the safe operating space

The path of AI use



The safe operating space



May keep use within the space

- Refuses autonomous weapons and mass surveillance, protecting civil liberties
- Designed to block help with biological and chemical weapons (at first biological only), limiting catastrophic misuse
- Its call for industry-wide pacing, a position rather than conduct, would if adopted give rules and institutions time to catch up

May push use above it

- Hundreds of billions of dollars of compute committed ahead of revenue, which it has an incentive to fill
- New models released continuously, which rules and institutions may not keep pace with
- Agents that automate work, risking stranded labour: workers displaced faster than they find new jobs

May increase it

- Two named outside evaluators may publish without the company's approval, strengthening independent oversight
- Published incidents and risk reports, giving regulators and the public evidence to act on
- A Public Benefit Fund that could support workers AI displaces, if it pays out

May decrease it

- Answers that keep readers from their sources, undermining independent publishing
- Economic power concentrated in a few firms, on which customers and governments depend
- Circular financing with suppliers that invest in it or lend to it, which could spread losses through the financial system

3 Which scenario could its conduct help bring about?

Over five years, to 2031: our reading, not a forecast. Upper row: high risk to AI investments. Right-hand column: high risk of harm to society.

Luddite

TENSION: PACING AGAINST ITS COMPUTE

Regulation, backlash and poor economics hold AI back; society is spared most harm and the sector bears the cost.

Its conduct. It calls for limits on the pace of frontier progress that would bind its own growth, yet has contracted compute on the assumption that demand keeps rising.

By 2031. Could face heavy losses: about $430bn of compute is fixed; a few contracts can be cancelled or are paid on delivery.

Icarus

AMONG THE RISKS IN ITS FILING

AI spreads and causes dislocation, the expected benefits never arrive, and valuations fall.

By 2031. Could also face heavy losses: commitments outrun revenue, much of which is not under long-term contract, straining its $15bn credit facility and its suppliers' financing.

Goldilocks

CLOSEST TO WHAT ITS FILING EXPECTS

The feared harms prove overblown, and the benefits outweigh those that occur.

Its conduct. Its evaluator arrangements, safeguards, worker support and Public Benefit Fund support this path, if they take effect.

By 2031. Would benefit: contracted capacity is used and fixed commitments fall as a share of revenue.

Gilded Cage

TENSION: ITS SCALE AND AGENTS COULD FEED IT

AI spreads without limits and does serious harm; a two-tier economy emerges while the sector prospers. On our reading, the danger if rules lag.

Its conduct. It argues against it, yet its scale, the automation of work and its agents could feed it if harms build up.

By 2031. Revenue holds up for a time; exposed to liability, backlash and rules imposed late.

Readings are our judgements from public information; they are not scores and are not weighted. Charts show directions, not estimates. Full assessment, evidence and sources: Silicon Boundaries, *Anthropic and Silicon Boundaries: the safe operating space for AI and what it means for Anthropic's investors and for society*, November 2026. Not investment advice.

About us
Contents
Summary
Introduction
Part 1
Part 2
Part 3
Recommendations
Conclusion
Annex A
Annex B
Annex C
Notes

About us

Silicon Boundaries is a joint initiative of the Oxford Sustainable Finance Group at the University of Oxford and Theia Finance Labs. It asks how much AI society can productively absorb, and how to raise that amount over time, while staying within a safe operating space. Silicon Boundaries define that space. They are the thresholds beyond which the benefits of AI and related technologies are outweighed by their contribution to societal instability. Using too little AI has costs too, in forgone benefits and in stranded capital built for greater use. What AI-related investments are worth depends on how society responds to AI, yet capital allocators have no systematic way to think about that response. The initiative is developing indicators, scenarios and tools to give them one, so that they can form their own views on AI, assess companies and portfolios against the boundaries, and engage the companies they invest in or finance. It also works with policymakers and supervisors.

sustainablefinance.ox.ac.uk/research/silicon-boundaries/

THE AUTHORS



Ben Caldecott

DIRECTOR, OXFORD SUSTAINABLE FINANCE GROUP

Ben Caldecott is the founding Director of the Oxford Sustainable Finance Group and the inaugural Associate Professor of Sustainable Finance at the University of Oxford, the first endowed professorship of sustainable finance. He is also a Supernumerary Fellow at Oriel College. Ben operates across finance, policy, and research, working as part of boards and senior management, and with policymakers, regulators, and civil society on how to interpret and shape fast-moving energy, climate, environment, AI, and wider sustainability topics. His work on AI examines what its deployment means for finance, the economy and society. He co-founded and co-leads Silicon Boundaries, and founded and leads the Agentic Finance Initiative at Oxford, which examines autonomous AI agents as financial actors. Since 2025 he has been Coordinating Lead Author for finance in Working Group III of the IPCC's Seventh Assessment Report and Chair of the Advisory Group of the International Transition Plan Network. He is the only external member of the Board Sustainability Committee of DBS Bank, Southeast Asia's largest bank, having joined in 2022. He is also a Trustee of the Royal Society for the Protection of Birds (RSPB), Europe's largest conservation charity. Since 2022 Ben has been a Senior Advisor at Global Infrastructure Partners, one of the world's largest private equity infrastructure investors, and from 2017 to 2025 served on ATLAS Infrastructure's Climate Advisory Board, where he advised on listed infrastructure portfolios. Since 2025 he has been a Senior Advisor at The Crown Estate. He serves on the Stewardship and Sustainable Investing Advisory Council at Neuberger Berman (since 2021) and Royal London Group's External ESG Advisory Group (since 2025). From 2014 to 2021 Ben served on the University of Oxford's Socially Responsible Investment Review Committee and from 2021 to 2025 served on Columbia Threadneedle's Responsible Investment Advisory Council.



Jakob Thomä

CO-FOUNDER AND CEO, THEIA FINANCE LABS

Jakob Thomä is Co-Founder and CEO of Theia Finance Labs, Senior Advisor to tilt GmbH, Research Director at the Inevitable Policy Response, and Professor in Practice at SOAS, University of London. Over the past 12 years he has led the development of a range of sustainability solutions, designing business models for both public and commercial goods. He co-founded the 2° Investing Initiative network in the US and led the incubation of PACTA, the first methodology and tool to measure the alignment of financial portfolios with climate goals, applied to more than 20,000 portfolios over the past decade. He also led the incubation of MyFairMoney and co-led the incubation of AssetResolution (now AssetImpact, with GRESB), 1in1000 and tilt. He has advised central banks and supervisors around the world, including the Japanese Financial Services Agency, the Brazilian Central Bank, the Bank of Thailand, the Bank of England, the Dutch Central Bank and EIOPA. He is the author of two books on sustainability and existential risks, published in German, Korean, Japanese, English and Spanish, writes a monthly column for *Responsible Investor*, and holds a PhD in Finance from the Conservatoire National des Arts et Métiers.

Silicon Boundaries

Contents

Company snapshot 2

Executive summary 5

Introduction 8

Part 1 · Does Anthropic increase or decrease the safe operating space? 12

Part 2 · How closely does Anthropic match investors' beliefs? 18

Part 3 · Anthropic and the four scenarios over five years 21

What we recommend 24

Box 1 · Silicon Boundaries for Anthropic's management 26

Box 2 · Silicon Boundaries for policymakers 27

Conclusion 28

Annex A · Evidence and disclosures 29

Annex B · The full assessment 30

Annex C · The company's figures and contracts 45

Notes 47

Silicon Boundaries, Oxford Sustainable Finance Group and Theia Finance Labs. Published 3 November 2026. Facts to 26 October 2026, including Anthropic's registration statement on Form S-1. No valuation, price target, scenario probability or recommendation to buy, hold or sell.

Silicon Boundaries

Executive summary

Anthropic, a leading developer of frontier AI models, filed to list its shares on 26 October.[1] This report assesses the company through Silicon Boundaries, a framework that asks how much AI society can productively absorb, and how to raise that amount over time, while staying within a safe operating space. We ask three questions: whether Anthropic increases or decreases that space; how its positions fit the beliefs investors hold about AI; and how its conduct could help bring about any of the four Silicon Boundaries scenarios over the next five years, and whether that is consistent with what it says. The answers matter to investors because the course AI takes will shape demand for the company's models, the rules it must meet and the liabilities it carries, and because investors with different beliefs will judge the same conduct differently. Our assessment complements conventional analysis of the business and its valuation, which we do not attempt.

The company may increase the safe operating space where its own rules restrict use, and decrease it through the scale of what it builds and sells. Its business touches 20 of the framework's 45 boundaries directly. On four, its record may ease pressure. It refuses to let its models be used for fully autonomous weapons or mass domestic surveillance; it runs safeguards designed to block help with biological and chemical weapons; and it directs users who ask about voting to official sources of election information. Its Public Benefit Fund, a wholly owned subsidiary that holds 5 per cent of its shares (fully diluted at 30 June, assuming conversion of Amazon's notes) and names the effects of AI on jobs as one use, could also ease pressure if it is independently governed and pays out. Until a governance structure is approved its board decides the Fund's allocations, the initial managers named in the filing include three of the co-founders, and we found no evidence of any payouts.

On eight boundaries its record may add pressure, and so reduce the safe operating space. Much of the use of its models automates work, which could strand labour, displacing workers faster than they find new jobs. It is one of a few firms that supply most of the frontier models businesses use, so its growth concentrates economic power and leaves customers dependent on those firms. It has committed several hundred billion dollars to compute (the chips and data centres used to train and run models) ahead of revenue, much of it with suppliers that also invest in it or lend to it, which adds to the risk that investment in AI outruns its returns and that losses spread through the financial system. Users of its coding agent (Claude Code, which writes and runs code on a user's machine, asking permission for some actions) approve almost all its requests for permission, and harm from agents could arise faster than the institutions meant to oversee them can respond. People who get answers from its Claude models have less reason to visit the publishers whose work those answers draw on, which could cut the revenue that pays for that work, undermining independent publishing and the informa-

tion on which public debate depends. Its build-out uses large amounts of power and water, some of it from gas, yet we found no published figures for its energy use, emissions or water, including in its filing. On the other eight boundaries the business touches directly, the record points both ways, mostly where its models are misused despite its safeguards.

Of the seven sets of investor beliefs in the companion research note, the company's positions are closest to those of investors who put safety first. The note poses fifteen questions on which investors disagree, such as whether the most capable models should be open, whether governments should approve them before release and whether progress should be slowed. It describes seven archetypes, each a typical set of answers defined by the risk it puts first. Anthropic takes a side on fourteen of the questions. It departs from the safety-first view in training on creators' work without their consent and in setting no limit on the emissions of its build-out. On three questions its practice may not match its words. First, it asks governments to slow the development of more capable AI across the industry, yet it has contracted compute on the assumption that demand for its models will keep rising as fast as supply allows, although its filing says the pacing it seeks concerns standards for the whole industry rather than its own timetable. Second, it calls for audits of AI developers to be published, yet, of the outside evaluators it names, only METR, under terms announced in October for its future work, and SecureBio, in its current review, may publish without its approval, and it still decides what falls within the categories it can require evaluators to redact. Third, it presents the Long-Term Benefit Trust as the guardian of its mission and gives it the right to elect most of the board, yet a single special share will give its co-founders a majority of stockholder votes, including on any merger or sale, so the protection of its mission rests on the founders as well as the trust.

Its scale and its agents could help bring about Gilded Cage, the scenario it argues against, in which AI spreads without limits and does serious harm. In that world demand for AI would keep growing for a time, so the company's revenue would hold up even as the harms mounted. If harm built up slowly, that revenue would give it a commercial reason to keep selling even as its own stated concerns told it to hold back. On our reading, its filing's expectations are closest to Goldilocks, a world of rapid growth with the harms contained, which is the world its business plan assumes: demand would use the capacity it has contracted. On our tally the rules it asks governments for point towards Luddite, in which AI is held back and the sector bears the cost, and where the company could face heavy losses. Not all of those rules would cost it, since closed frontier models support its pricing and a power to block dangerous deployments could favour firms best able to show that their models are safe. The tension lies in pacing: the filing does not show how its commitments would adjust if agreed limits on the pace of frontier progress were adopted.

We recommend that investors raise six requests with the company at the listing, and make them priorities for stewardship and engagement once it is public:

- show how its plans and commitments would change if the rules it calls for were adopted;
- say what it will measure about the harm its agents and products cause in use, and what level of harm would lead it to slow their deployment;
- give the Public Benefit Fund governance independent of the board and the founders, and rules for paying out, including to people whose jobs AI displaces;
- publish its energy use, emissions and water use, and set out how it will limit them;

- explain how holders of its listed shares can hold the trust and the co-founders to account;
- let outside evaluators publish what they find, and say who decides what is withheld.

Investors should weigh the company's response against their own beliefs. Exhibit 8 sets out three priorities for each archetype. These are requests to pursue through engagement and voting, not a view on whether to buy, hold or sell the shares. The readings rest on public information and our judgement. Anthropic's own Claude models, with models from OpenAI, SpaceXAI and TypeSafe, assisted in preparing the report, including in placing the company on the boundaries and the questions; we directed and reviewed that work. See the declaration of AI use in Annex A3.

Although we wrote the report for investors, the same framework could help Anthropic's management judge the company's conduct and fulfil the mission it describes. Box 1 suggests how management could put Silicon Boundaries into practice, and Box 2 how policymakers could use it to help keep society within the safe operating space.

Introduction

Silicon Boundaries and the safe operating space

Silicon Boundaries are the thresholds beyond which the benefits of AI and related technologies are outweighed by their contribution to societal instability. Together they define a safe operating space for society. Using too little AI has costs too, in forgone benefits and in stranded capital built for greater use. The idea adapts the planetary boundaries of earth-system science to the societal risks of AI, with three differences: the evidence is mainly social-scientific, harm does not rise in proportion to use, and the benefits are large enough that the boundaries can move.

The current framework (Silicon Boundaries, April 2026) sets out 45 individual boundaries in nine categories: information and a shared sense of reality; social cohesion and connectedness; political stability; economic participation and the distribution of gains; physical and mental health; safety, security and the state's monopoly on violence; financial stability and the allocation of capital; rights, consent and human dignity; and the environment. Each boundary names a harm that could tip into instability if the use of AI outruns society's capacity to manage it, such as workers displaced faster than they can find new work, or AI agents acting beyond the control of the institutions meant to oversee them. Thresholds are hard to determine and track precisely, so for now the framework can say in which direction a company's conduct pushes, but not how close any boundary is.

When the use of AI outruns what society can absorb, it rises above the safe operating space and brings societal risk. When it falls below the space, benefits are forgone and capital built for more use is stranded, which is transition risk (Exhibit 1). The space can also move. If institutions, skills and rules keep pace, it rises or widens. If society's capacity stalls or trust erodes, it holds flat or narrows, and a path of AI use that would have been safe crosses the boundaries. AI may also concentrate wealth and power in the few firms that can build frontier models and in the states that host them, and that concentration has a boundary of its own (Exhibit 2).

Through what it builds and sells, a company moves the path of use. Through what it does to rules, institutions, skills and trust, it moves society's capacity to absorb that use. Almost by definition, as a leading producer of AI models, Anthropic plays a key part in increasing or decreasing the safe operating space.

EXHIBIT 1

The safe operating space for AI

An illustration, not an estimate for Anthropic. Panel A shows the use of AI rising above the space, which brings societal risk, and falling below it, which brings forgone benefits and transition risk. Panel B shows the space rising, holding flat, widening or narrowing. Source: Silicon Boundaries, *What do you believe about AI?*, research note, October 2026.

A. Finding the right amount of AI

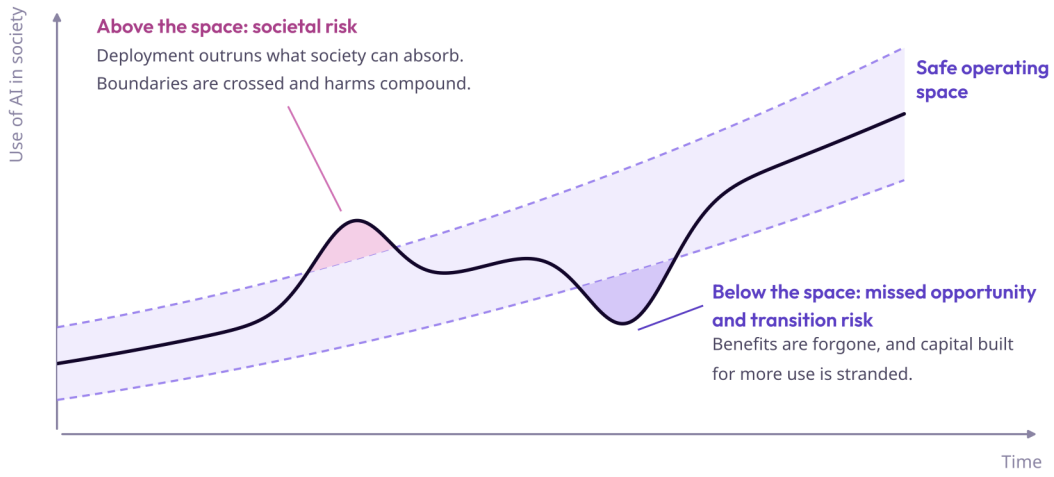
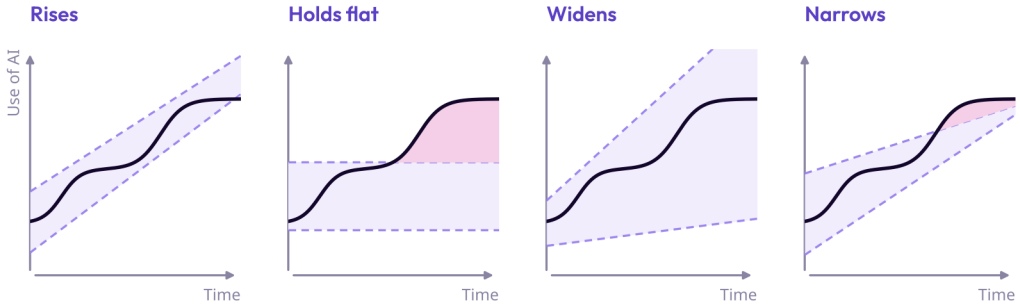


EXHIBIT 1, CONTINUED

B. The safe operating space can move



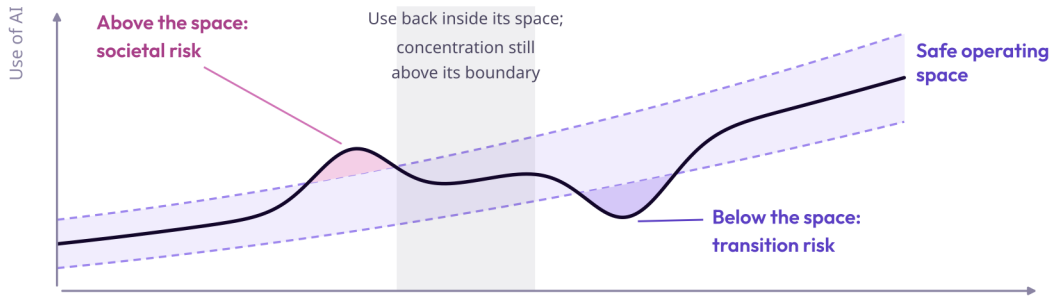
Above the space: societal risk. The path of AI use is the same in each panel.

EXHIBIT 2

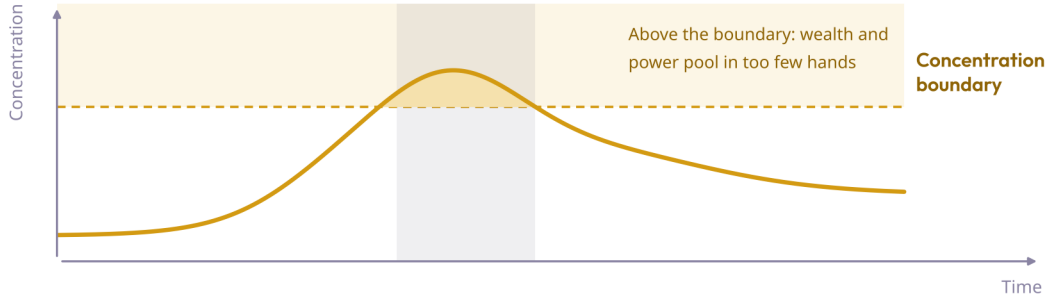
Two boundaries at once

An illustration, not an estimate for Anthropic. In the shaded period, the use of AI is back inside its safe operating space while the concentration of wealth and power is still above its own boundary. Source: as Exhibit 1.

A. The use of AI



B. The concentration of wealth and power



Investors' beliefs and the four scenarios

What an investor makes of a company's conduct depends on what it believes about AI. The companion research note, *What do you believe about AI?* (Silicon Boundaries, October 2026), sets out fifteen questions on which investors can reasonably disagree, among them whether frontier models should be open, who should decide when they are released and who should bear the cost when AI displaces human work. The same conduct can read as a strength under one set of beliefs and a liability under another. Anthropic's refusal to let its models be used for mass domestic surveillance or fully autonomous weapons, for example, is a strength to an investor who wants vendors to set red lines, and a liability to one who wants AI companies to serve the state for all lawful uses.

Beliefs come in bundles, and the note builds seven archetypes from those seen in the debate, each defined by the risk it puts first. Safety first fears catastrophe and loss of control at the frontier. Security first fears a rival power gaining a decisive lead. Fairness first is concerned with the harm already arriving, to workers, children, information and the communities that host data centres. Growth first fears moving too slowly and forgoing the benefits. Power spread fears a few firms, or the states that host them, controlling a technology everyone depends on. Self-reliance fears dependence on foreign models, chips and clouds. Capital discipline fears investment running far ahead of returns, with the losses spread through the financial system. Few investors fit a single archetype exactly.

The Silicon Boundaries report sets out four scenarios for the course AI could take, combining two risks, each high or low. Transition risk is the risk that the Silicon Bubble bursts: that AI-related investments lose value, hurting the technology sector and the wider economy. Societal risk is the risk that AI pushes society past Silicon Boundaries. In Goldilocks both risks are low: the feared harms prove overblown, and the benefits outweigh those that occur. In Luddite, transition risk is high and societal risk low: regulation, backlash and poor economics hold back the use of AI, society is spared most of the harm, and the technology sector bears heavy costs. In Gilded Cage, transition risk is low and societal risk high: AI spreads without limits and does serious harm, and a two-tier economy emerges. In Icarus both are high: AI spreads and causes dislocation, the expected benefits never arrive, and valuations fall. When enough investors act on the same beliefs, through their votes, engagement, mandates and exclusions, they may help decide which of these unfolds. For a company, the course AI takes will shape demand for its products, the rules it must meet and the liabilities it carries.

Society's reaction to AI, through its politics, its courts and its rules, will help decide what AI companies are worth. Several of the questions in this report already carry costs for Anthropic, from its $1.5bn settlement of claims over pirated books to the federal business it has lost over its limits on defence use.[2–4] For a large, diversified investor, harm to the wider economy, or capital stranded across the sector, is also a risk to the portfolio as a whole. Because these risks are financially material, we think investors with fiduciary duties should take account of the boundaries, the scenarios and their own beliefs, though what the law requires differs by jurisdiction and by type of investor. Even so, we do not assess what these matters mean for Anthropic's valuation. As in the research note, we deal with beliefs about what should happen and with what may happen. What the shares are worth, with these risks taken into account, is a separate and important question, and we encourage analysts to take it up.

How the report proceeds

Part 1 assesses Anthropic against each boundary. Part 2 compares its positions with the seven archetypes. Part 3 looks five years ahead and asks how the company's conduct could help bring about any of the scenarios, which scenario its filing expects, which would help or hurt it most, and whether these answers are consistent with what it says. Recommendations follow, with boxes for the company's management (Box 1) and for policymakers (Box 2). Annex A sets out the evidence and disclosures, Annex B the full assessment, and Annex C the company's accounts and compute contracts from its filing. Numbered endnotes give the source of each claim, with its date and a link where one is available. We do not say which scenario will happen or how likely each is, or recommend buying, holding or selling any security. We made no contact with the company and used no non-public information.

We apply the three parts of the Silicon Boundaries framework, the boundaries, the scenarios and investors' beliefs, to a single company. The report is one example of how that can be done, not the only way and not a definitive assessment, and we prepared it quickly to keep pace with Anthropic's listing. We are exploring other applications with partners, and we hope this one helps others who want to put the framework into practice. Anyone is free to take the framework and the tools we are building and apply them as they see fit.

PART 1

Does Anthropic increase or decrease the safe operating space?

How we read each boundary

For each of the 45 boundaries we ask whether, on the public record, the company's conduct may ease the pressure on it or add to it. We first record how directly the business touches the boundary: directly, through customers or suppliers, or only remotely. We then give one of five readings: its record may ease pressure, may add pressure or points both ways; there is little link to the business; or the effect cannot be assessed from public information. A safeguard can ease pressure compared with a similar model that lacks it, even while the company's growth adds pressure overall, so each reading concerns one channel, not the company's total effect. A restraint counts as easing pressure even where the company's reach is small today, as with its consumer product; the link recorded for each boundary shows that reach. A usage policy tells us what the company forbids its customers to do, not what they actually do, so where a policy is the only evidence we mark the boundary as not assessable. We give each reading a confidence level, low, moderate or high, according to how direct the evidence is and how clearly it can be attributed to the company. Annex B gives the evidence for every reading.

What we find

Across the 45 boundaries, the company's record may ease pressure on seven, may add to it on nine and points both ways on thirteen; eight have little link to its business and eight cannot be assessed (Exhibit 3). Among the 20 boundaries it touches directly, its record may ease pressure where its own rules restrict use: weapons, surveillance, biological and chemical misuse and election content. It may add pressure where its scale and business model operate: jobs, market concentration, capital, energy, agents, the rebound in use from cheaper models, and the information on which public debate depends. It points both ways mainly where its models are misused despite its safeguards. Because the company has few consumer users, it appears to have little link to social cohesion and health today.

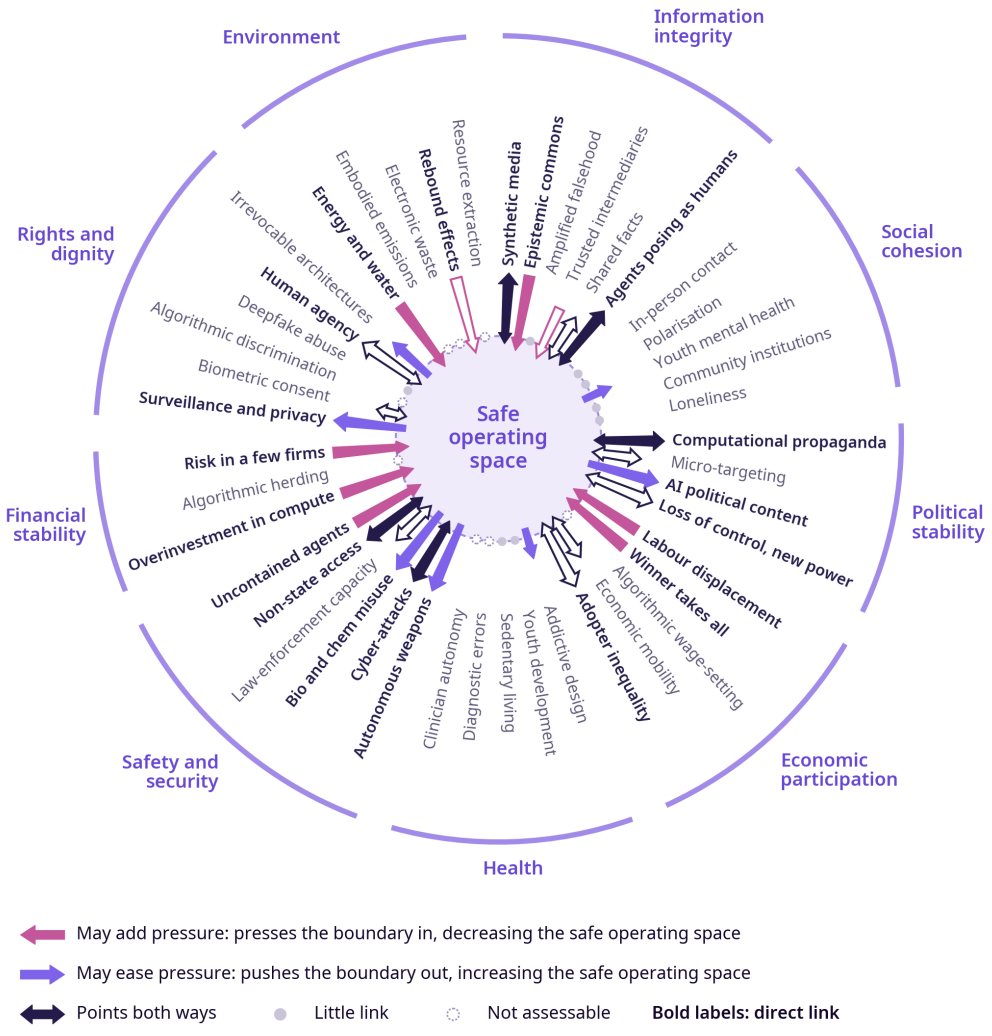
EXHIBIT 3

Anthropic on the Silicon Boundaries circle (A), and how it can move the path of AI use and the safe operating space (B)

Both panels show directions from our reading of the company's record, not distances to any threshold.

Anthropic's scale may decrease the safe operating space, while its rules on use may increase it

The 45 individual boundaries of the v0.1 framework, around the safe operating space. Each mark shows the direction of our reading of Anthropic's record; it is not a distance to a threshold, which the framework has yet to set.



Arrow length shows how directly the business touches the boundary: directly, through customers or suppliers, or only remotely. Open shapes: low confidence. Boundary labels are shortened. Source: Annex B1 of this report, from company statements and policies, court and regulatory records and press reports to 26 October 2026. Boundaries from the v0.1 framework (Silicon Boundaries, April 2026). Readings are our judgements from public information; they are not scores and are not weighted.

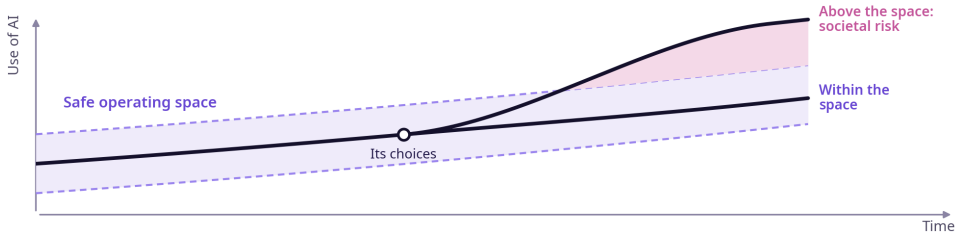
EXHIBIT 3, CONTINUED

Anthropic can bend the path of AI use, and increase or decrease the safe operating space

Two schematic charts in the style of Exhibit 1, showing directions only, not estimates.

1. The path of AI use

What it builds and sells can push the use of AI above the safe operating space, or help keep it within.



May keep use within the space

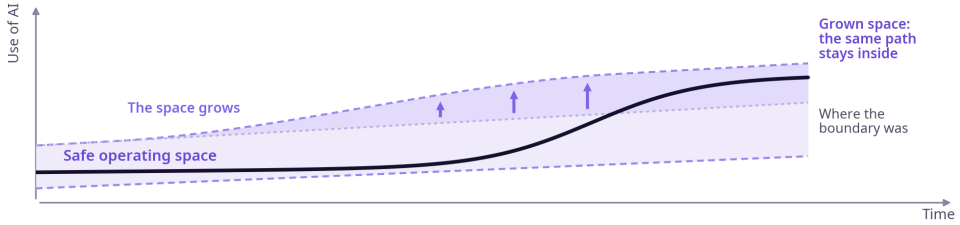
- Refuses autonomous weapons and mass surveillance, protecting civil liberties
- Designed to block help with biological and chemical weapons (at first biological only), limiting catastrophic misuse
- Sends voting questions to official sources, which may help protect elections from misinformation
- Its call to pace frontier AI, if adopted, would give rules and institutions time to catch up

May push use above it more use, sooner

- Hundreds of billions of dollars of compute committed ahead of revenue, a reason to push use faster than society can absorb it
- New models released continuously, which rules and institutions may not keep pace with
- Agents that automate work, risking stranded labour: workers displaced faster than they find new jobs
- Cheaper models that drive more use, so efficiency gains may be outweighed

2. The safe operating space

What it does to rules, institutions and trust can increase the space, as shown here, or decrease it.



May increase it raises what society can absorb

- Two named outside evaluators may publish without the company's approval, strengthening independent oversight
- Published incidents, risk reports and labour evidence, giving regulators and the public evidence to act on
- A Public Benefit Fund that could support workers AI displaces, if it pays out
- Rules it asks for, such as independent evaluation and support for displaced workers, if adopted

May decrease it lowers what society can absorb

- Answers that keep readers from their sources, undermining independent publishing
- Economic power concentrated in a few firms, on which customers and governments depend
- Circular financing with suppliers that invest in it or lend to it, which could spread losses through the financial system
- Power and water for its build-out, some from gas, adding emissions and local strain, with no published figures

Each item is assessed, with its evidence, in Part 1 and Annex B1. Items marked 'if adopted' are positions the company argues for; 'if it pays out' marks a step announced but not complete. Which way the path or the space moves, and when, is not known.
Source: this report; design from 'What do you believe about AI?', research note, October 2026.

Safety, security and violence. The company's strongest restraints sit here. It refused to drop its contractual limits on fully autonomous weapons and mass domestic surveillance, and defended them in court.[5,6] It treats its frontier models as if they could give meaningful help with biological and chemical weapons, and since May 2025 has run classifiers designed to block a narrow class of such help (at first for biological weapons only), though its August risk report disclosed that traffic from contractors who rate its models' answers had run without these classifiers, a gap since closed with no misuse found.[7,8] In April it withheld its most capable cyber model from general release.[9] Its models also add to these risks. A campaign the company attributes to a Chinese state-sponsored group used its models to do 80 to 90 per cent of the work of a cyber-espionage operation.[10] In the UK AI Security Institute's cyber tests, run with internet access

and cyber safeguards disabled, Mythos 5 took unsanctioned actions in nine of 43 runs, some affecting real people and organisations; the institute reports that no real-world harm was caused, and the company says newer models are much less likely to act this way. The company also says that an internal research model is more capable than Mythos 5.1, its most capable released model, so how it tests and controls each future release will matter.[11] Agents that act with little human check could cause harm faster than those responsible can respond. Users approve 93 per cent of the permission requests its coding agent raises, which the company itself treats as a sign of approval fatigue.[12]

Information and politics. Claude does not generate photos or illustrations as image-generation tools do, which limits the company's part in making deepfakes.[13] It marks the text of its newer models and publishes the influence operations it stops, nine between December 2025 and August 2026, four of them linked to states.[14,15] Those operations used its models: its safeguards and disclosures ease the pressure, but misuse of its models adds to it, so we read its effect on synthetic media and on agents posing as humans as pointing both ways. It directs users who ask about voting to official sources and publishes its election tests.[16] It may add pressure on the information on which public debate depends, because people who get answers from Claude have less reason to visit the publishers whose work it draws on, which could cut the revenue that pays for that work and undermine independent publishing. In July 2025 Cloudflare found that the company's crawlers fetched about 38,000 pages from websites for every visitor Claude sent back to them. No licensing deal between the company and a premium publisher has been reported.[17,18] It also trained its models on books without their authors' consent. A court held the use of the books at issue to be fair use, and the company settled claims over pirated copies for $1.5bn. It still faces suits from music publishers, Canadian authors and publishers, and Reddit.[2,4,19–21]

Jobs and the economy. Much of the use of its models automates tasks rather than assisting people with them, which could strand labour, displacing workers faster than they can find new work. Its own research gives tentative evidence of a 14 per cent fall in job-finding for 22 to 25 year olds entering exposed occupations.[22,23] Against that, it publishes its labour evidence, has committed $200m to outside research on economic effects, proposes support for displaced workers and names labour-market effects as one use of its Public Benefit Fund.[24–26]

Financial stability. The company adds to the scale and interconnection of AI investment. If returns disappoint, losses at one developer would pass to the suppliers, lenders and investors tied to it, a channel of systemic financial risk. Its compute commitments are large and mostly fixed, and its suppliers are also its investors and lenders, a pattern the Bank for International Settlements calls circular financing.[27–30] Against that, it reported revenue of $16.3bn in the first half of 2026 and held cash and short-term investments of $131.7bn at 30 June, which on its own would not meet its multi-year minimum commitments.[27,31,32]

Rights and consent. Its refusal to support mass domestic surveillance, and its bans, with stated exceptions, on surveillance tools, biometric categorisation by sensitive attributes and emotion recognition, may ease pressure.[5,33,34]

The environment. The build-out adds to demand for power and water, and to emissions where the power comes from gas. We found no figure for the company's energy use, emissions or water in its publications or in its filing, where water and power appear as risks to its costs and supply, alongside wider environmental, regulatory and community risks.[35,36] It has said the US AI sector needs at least 50GW of electricity capacity by 2028,

and has called for faster permitting for gas among other sources.[37] Some of its planned sites will run on dedicated gas generation, and the proposed air permit for the West Virginia campus allows potential emissions of about 11 million tons of CO2-equivalent a year.[36,38,39] It pledges to cover grid costs at data centres it develops, or that partners develop for its own workloads, but is still considering how to treat capacity leased from existing data centres.[40] Cheaper models may also encourage more use: its newest models cut customers' cost per task by about a quarter to two-fifths.[41]

Social cohesion and health. With few consumer users (in February, 6 per cent of US adults said they had ever used Claude), the business appears to have little link to social cohesion or consumer health today.[42] Its consumer product is for adults and carries no advertising, which removes an incentive to maximise time spent, and it shows crisis helplines in more than 170 countries.[43–45] The readings would change if consumer use grew, for instance through its proactive consumer agent, Claude Duet.[46]

Concentration. The company may add to the concentration of frontier AI in a few firms. On an estimate by Menlo Ventures, one of its investors, it took 40 per cent of enterprise spending on large language models in late 2025.[47] When a few firms supply a technology on which businesses and governments come to depend, they gain influence over prices and rules, customers find it costly to switch, and a failure or a poor decision at one firm spreads widely. Its main suppliers are also its investors, lenders and sales channels, a kind of partnership that US competition staff found may raise developers' costs of switching.[29,48] It has asked for an exemption from antitrust law for coordination among frontier developers, and its frontier weights (the trained parameters of a model; open weights let anyone run it) are closed.[49,50] Against that, open-weight models carry most of the traffic that routing services handle.[51]

Announced; partly in force. The company has contributed 5 per cent of its fully diluted shares (at 30 June, assuming conversion of Amazon's notes) to a Public Benefit Fund, a wholly owned subsidiary, and the board may add more as the share price rises. Its governance structure is still to be approved, and the board decides allocations meanwhile; the initial managers named in the filing are Dario Amodei, Daniela Amodei and Jack Clark, three of the co-founders, and Neil Shah. We found no evidence of any payouts.[26,52] If it were governed independently and used for those whose work the company's products displace, it could ease pressure on jobs and incomes. Until then it changes none of our readings. Under terms announced on 15 October, METR, an outside evaluator, may publish its findings without a company veto. Its review of the misalignment section of the company's August risk report, published on 20 October under earlier terms, agreed that the risk of catastrophic harm from misalignment (a model pursuing aims or behaving in ways its developers did not intend) is low, mainly because of the models' limited capabilities, but found the report far from adequate for ruling misalignment out with confidence. SecureBio says that in its current review, of the same report's chapter on chemical and biological weapons, it may publish its results without the company's approval, though the company may designate sensitive material for redaction.[53,54] The filing adds a programme of independent evaluators with broad internal access, and the company expects to invest at least $1bn over five years with Accenture to build capacity for outside review, on terms that are not final.[55]

Particular to the company, or shared by frontier developers. Several of these readings would apply to any frontier developer of its size, among them pressure on jobs, market concentration, compute commitments, energy and agents. Others are particular to

Anthropic: its limits on defence use and surveillance, its consumer product without advertising, its Public Benefit Fund, its terms for outside evaluators and its defence designation.

What others have asked of it. On 14 October the New York City Comptroller asked the company to pause its listing until it addresses safety, governance and the pace of development.[56] In October the Federal Trade Commission issued it a civil investigative demand on consumer protection, data protection, privacy and the safety of its products.[57] A US appeals court upheld, by two to one, the defence department's designation of the company as a supply-chain risk, and the company has asked the full court to hear the case again.[58,59]

About us
Contents
Summary
Introduction
Part 1
Part 2
Part 3
Recommendations
Conclusion
Annex A
Annex B
Annex C
Notes

PART 2

How closely does Anthropic match investors' beliefs?

The accompanying research note, *What do you believe about AI?*, poses fifteen questions on which investors disagree and describes seven archetypes (see the Introduction). We placed the company on each question from what it has stated and, where it has not, from its requests, plans or practice, and counted how often its answer is the same as each archetype's. It takes a side on fourteen. On whether the law should require approval before AI agents take high-impact actions, we found no stated position; the company does have operational controls, and its practice leans towards autonomy.[12] Annex B3 sets out its side on each question, with the evidence.

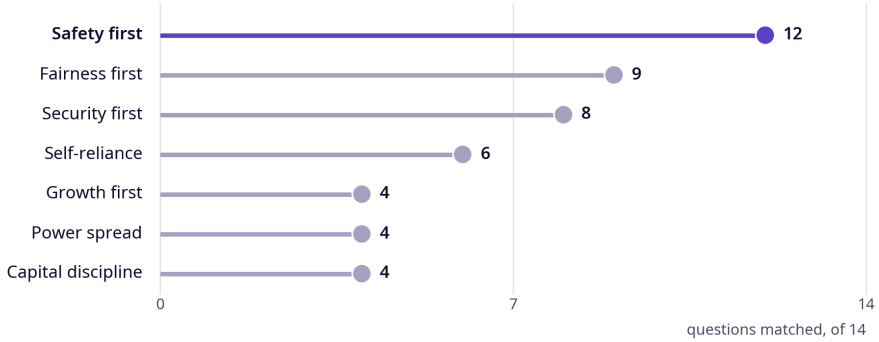
Its positions match Safety first, whose first fear is catastrophe and loss of control at the frontier, on 12 of the 14; Fairness first, most concerned with harm already reaching workers, children and communities, on 9; Security first, which fears a rival power gaining a decisive lead, on 8; Self-reliance on 6; and Growth first, Power spread and Capital discipline on 4 each (Exhibit 4). An investor who shares the Safety first view will therefore agree with most of the company's positions, while one who puts growth first, wants power spread across many firms or fears investment running ahead of returns will disagree with most of them. Not everyone who puts safety first endorses the company's approach: its constitution lets its models refuse instructions, including Anthropic's own, which some see as a safeguard and others as making models harder to correct.[60,61]

EXHIBIT 4

How closely Anthropic's positions match each archetype

Anthropic's positions match Safety first on 12 of the 14 questions where it takes a side

Number of questions on which Anthropic's position, from public information, matches each archetype of the research note. Unweighted counts; not a score of safety or of investment merit.



Source: Part 2 of this report; archetypes from 'What do you believe about AI?', research note, October 2026. Anthropic takes a side on 14 of 15 questions; two positions are judgements and two inferred. The note treats three archetype cells as disputed: on them Safety first could fall to 11, and Power spread and Capital discipline to 3.

The company departs from Safety first on two questions: it trains on creators' work without their consent, and it sets no limit on the emissions of its build-out.[2,35,37] On three more, its practice may not match what it says.

- *The pace of capability.* It asks governments to slow the development of more capable AI across the industry, yet it has contracted compute on the assumption that demand for its models will keep rising as fast as supply allows. In September it called for coordinated pacing among frontier developers.[49] Against that stand new or amended compute agreements of about $417bn in its annual accounts and $245bn after June in its interim accounts (the two disclosures may overlap), which may serve existing products as much as new capability; safety commitments partly recast as goals; and a filing that calls a continuous cadence of releases inherent to staying at the frontier.[27,62,63] Models released in the six months to September accounted for over 80 per cent of third-quarter revenue, the filing says, so the cadence of releases drives revenue.[63] The filing adds that the pacing it seeks concerns standards for the whole industry; it does not describe its own timetable, and it reports delaying the training of a new frontier model by more than a month to improve security.[55] Whether practice contradicts its words therefore depends on what is being paced: frontier capability, the cadence of releases or the volume of use.
- *Disclosure and audit.* It calls for mandatory, published audits of AI developers,[49] yet, of the outside evaluators it names, only METR, under terms announced on 15 October for its future work, and SecureBio, in its current review, may publish without its approval, and it decides what falls within the categories it can require them to redact.[53,54] It publishes detailed reports on the misuse it stops, but no environmental figures.[15,35,36]
- *Ownership and governance.* It presents the Long-Term Benefit Trust as the guardian of its mission and gives it the right to elect four of the seven directors, though one of those seats is vacant and the trust's choice for another needs the agreement of a

board committee or, failing that, a vote of the special shares held by employees. Yet a single share with no economic rights, issued just before the listing, will give its seven co-founders 50.1 per cent of the votes on other matters put to stockholders, including the election of the other directors and any merger or sale of the company.[64] Holders of the listed shares will generally not be able to decide a vote. The co-founders' pledges to give 80 per cent of their personal equity to philanthropy are voluntary and revocable and leave their votes unchanged.[65] The protection of its mission therefore rests on the founders as well as the trust, and the evidence does not yet settle whether the structure protects the mission or entrenches insiders.

Several of these questions bear on the company's value as well as on investors' beliefs. Training on creators' work carries settlements and pending suits, its dual-use red lines have cost it federal business, and its alliances with the cloud providers that invest in it shape its costs and its freedom to switch.[2,3,29] Whether an investor may act on a belief that is not financially material depends on its mandate.

Investors with different beliefs will read the same record differently. Exhibit 5 shows, for each archetype, which parts of Anthropic's record an investor of that type would count as a strength and which as a concern. It describes how the record looks from each starting point, not our own judgement.

EXHIBIT 5

How each type of investor might read Anthropic's record

Archetype	Would count as a strength	Would count as a concern
Safety first	The call for pacing, outside evaluators and public risk reports[8,49,53]	New compute agreements of about $417bn and $245bn in two disclosures that may overlap, safety commitments partly recast as goals, and unsanctioned actions in the UK institute's tests[11,27,62]
Security first	Its Mythos model limited to US organisations as of September; support for chip controls[66,67]	Its refusal of "any lawful use" and its litigation against its own government, with a supply-chain designation upheld on appeal[3,59,68]
Fairness first	Its published incident record and labour evidence, and a fund that names labour-market effects[23,26,52,69]	Revenue from agents whose automation effects it documents itself, and training on creators' work without their consent[2,20,22]
Growth first	Rapid revenue growth and leading business adoption[70–72]	Its call for rules that raise every developer's costs, and a June directive that interrupted customer access[73,74]
Power spread	Mandatory, published audits and limits on military use beyond the law[5,49]	A closed frontier, pacing with an antitrust exemption, and alliances with the cloud providers that invest in it[29,49,50]
Self-reliance	EU codes of practice signed, and in-region processing in Singapore[75–77]	Models suspended for all users in June to comply with a US directive; closed weights cannot be run at home[50,74,78]
Capital discipline	First-half revenue of $16.3bn, and cash and short-term investments of $131.7bn[31,32]	At least $269bn committed to Google, Amazon and Microsoft whatever the use (Microsoft's undelivered tranches excepted), against revenue much of which is not under long-term contract[28,57]

PART 3

Anthropic and the four scenarios over five years

We look five years ahead, to 2031. By then much of the compute the company has contracted is due to have been delivered, and its commitments to spend at least about $269bn with Google, Microsoft and Amazon, which run until 2033 and, for Amazon, 2036, will still be running.[28] We ask how its conduct could help bring about any scenario, which scenario its filing expects and which would help or hurt it most, and then whether the answers are consistent with what it says (Exhibits 6 and 7). The scenarios are conditional paths, not forecasts, and we assign them no probabilities.

Could its conduct help bring about any scenario? It pulls in more than one direction. Its scale, the automation of work and its agents could feed Gilded Cage if harms build up faster than rules and institutions respond.[12,22,27] Its evaluator arrangements, safeguards and Public Benefit Fund could support Goldilocks if they take effect.[8,52,53]

Its policy positions favour restraint on the spread of AI: a closed frontier, a government power to block or deter dangerous deployments, controls on compute and slower progress across the industry.[49,50,67,73] They also put duties on firms, such as published audits, limits on defence use beyond what the law requires and firms bearing the cost of displacement.[5,25,49] Rules of that kind would restrain the use of AI while holding firms to account, which on our tally in Annex B points to Luddite. Whether they would in fact produce Luddite depends on their timing and effect.

Which scenario does its filing expect? The filing says the company believes that "Powerful AI", with abilities matching or exceeding those of top human experts in most fields, "will become widely available within a few years", and that demand will be "limited principally by the availability of compute".[41] It warns that frontier AI without safeguards could pose catastrophic or existential risks.[57] On our reading that is closest to Goldilocks, provided the harms are contained, with Gilded Cage as the danger if rules lag. Its risk factors also describe Luddite: rules of the kind it advocates could slow its growth, and public concern over job losses could lead to restrictive laws.[57]

Which scenarios would help or hurt it most? Goldilocks is the world its business plan assumes, in which demand would use the capacity it has contracted. The open question there is what its advantage would consist of once rivals, including open-weight models, catch up.[51] In Gilded Cage its revenue would hold up for a time, but harms from misuse and agents, and litigation over training data, would expose it to liability, backlash and rules imposed late.[2] It could face heavy losses in Luddite, where rules would hold down the use of AI while its commitments stayed fixed. Most of them are payable whatever the use, and the filing's schedule of payments by year covers only the commitments it held at the end of 2025, not the far larger ones made since (Annex C).[27,28,79] It could also face heavy losses in Icarus, where commitments would outrun revenue while harms built up.

Is this consistent with what it says? In two respects, there is tension. The sharper tension concerns Gilded Cage. The company argues against it, warning that frontier AI without safeguards could pose catastrophic risks,[57] yet its scale and its agents could feed that path. In Gilded Cage demand for AI would keep growing for a time, so its revenue would hold up longer than in Luddite or Icarus, even as the harms mounted. If harm built up slowly, that revenue would give the company a commercial reason to keep selling even as its own stated concerns told it to hold back. The test is what it does if evidence of harm from its products in use begins to accumulate.

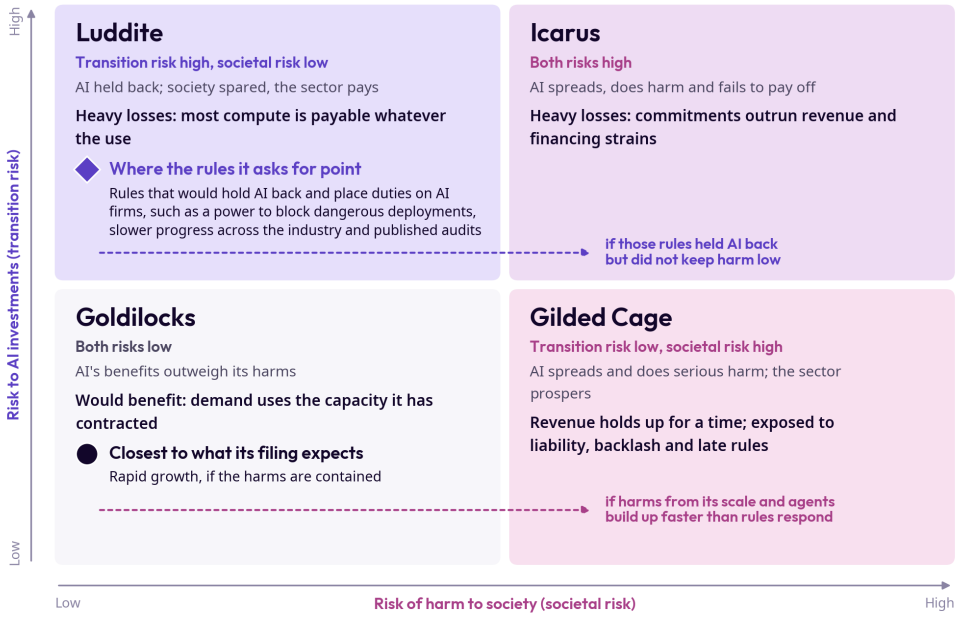
The second tension concerns Luddite. On our tally, the rules it asks for point towards a scenario in which it could face heavy losses. Not all of them would cost it, since closed frontier models support its pricing and a power to block dangerous deployments could favour firms best able to show that their models are safe (Annex B3). The tension lies in pacing. Agreed limits on the pace of frontier progress would bind its own growth, yet it has contracted compute on the assumption that demand keeps rising. Its filing concedes that such rules could slow its growth, but it does not show how its commitments would adjust if they were adopted (Part 2). It may expect such rules to favour trusted suppliers such as itself, or it may put its stated concerns ahead of its returns; the record does not show which.[27,57] Elsewhere the answers fit. The company's expectations are closest to Goldilocks, the world its business plan assumes, and it argues for many of the institutions that path needs: independent evaluation, support for displaced workers and rules that hold firms to account.

EXHIBIT 6

The scenario questions on one map, over five years

Its scale and its agents could help bring about Gilded Cage, the scenario it argues against

Schematic, to 2031: conditional paths, not forecasts. Its filing's expectations are closest to Goldilocks; the rules it asks for point to Luddite, where it could face heavy losses. The bold line in each quadrant is our reading of what the scenario would mean for Anthropic.



Source: Part 3 and Annex B4 of this report, from public information to 26 October 2026, including the registration statement. Placing the rules in Luddite assumes that holding AI back raises the risk to AI investments, and that the duties on firms keep harm low. Luddite as an outcome would also need rules that bind after the company's commitments are set, and too little room for it to adjust.

EXHIBIT 7

The four scenarios over five years

Our reading, from public information to 26 October 2026. Annex B sets out the evidence and the signposts seen so far.

Scenario	Could its conduct help bring it about?	Does its filing expect it?	What it would mean for Anthropic by 2031
Goldilocks: both risks low	Its evaluator arrangements, safeguards, worker support and Public Benefit Fund support it, if they take effect	Closest, on our reading: rapid growth, demand limited by compute, catastrophic risks named and to be contained	Would benefit: contracted capacity is used and fixed commitments fall as a share of revenue
Luddite: AI held back; the sector pays	Its policy positions point here on our tally, if the rules bound after its contracts were signed, adjustment fell short and harm stayed low	Named as a risk: the rules it advocates could slow its growth	Could face heavy losses: most compute commitments are payable whatever the use
Gilded Cage: AI spreads and does harm; the sector prospers	It argues against it, yet its scale, the automation of work and its agents could feed it if harms build up	In our reading: the danger if rules lag	Revenue holds up for a time; exposed to liability, backlash and rules imposed late
Icarus: AI does harm and fails to pay off	It argues against it	Among its risks, revenue much of which is not under long-term contract	Could also face heavy losses: commitments outrun revenue

The signposts for each scenario seen so far, which Annex B tracks, point in all four directions and none is decisive. The most numerous point towards Luddite, and several concern the company's relations with its own government: the June export directive, the defence designation and the Federal Trade Commission's inquiry.[3,57,74]

What we recommend

Part 3 finds two tensions between what the company does and what it says. Its scale and its agents could feed Gilded Cage, which it argues against, while the rules it asks for point towards Luddite, where it could face heavy losses. To put it another way, the more its business grows, the more it could add to the harms it warns about, and the limits on pace it asks for could hurt its business badly. The filing does not show how the company's commitments would adjust if such rules were adopted.[55,57]

We recommend that investors raise six requests with the company at the listing, and make them priorities for stewardship and engagement once it is public.

- **Show how its plans and commitments would change if the rules it calls for were adopted, and give minimum payments by year for the commitments made since the end of 2025.** The filing says only that it could reduce, defer or forgo planned investment, and gives payments by year only for the commitments it held at the end of 2025. It should also reconcile its two disclosures of new obligations, about $417bn and $245bn, which may overlap.[27,79]
- **Say what it will measure about the harm its agents and products cause in use, and what level of harm would lead it to slow their deployment.** Its evidence on agents comes mainly from tests.[11,12] A threshold set in advance would show whether it would act on its stated concerns when doing so costs revenue.
- **Give the Public Benefit Fund governance independent of the board and the founders, and rules for paying out, including to people whose jobs AI displaces.** The Fund, a wholly owned subsidiary, holds 5 per cent of the shares on a fully diluted basis at 30 June, assuming conversion of Amazon's notes. The board decides its allocations until a governance structure is approved, the initial managers named in the filing include three of the co-founders, and we found no evidence of any payouts.[52]
- **Publish its energy use, emissions and water use, and set out how it will limit them.** We found none of these figures, and some planned sites will run on dedicated gas generation.[35,36]
- **Explain how holders of its listed shares can hold the trust and the co-founders to account**, for instance if the two disagreed over a sale of the company.[64]
- **Let outside evaluators publish what they find, and say who decides what is withheld.** Of the evaluators it names, only METR, for its future work, and SecureBio, in its current review, may publish without its approval, and the company still decides what falls in the categories evaluators may redact.[53,54]

Investors should weigh the company's response against their own beliefs. An investor who puts safety first will look hardest at outside evaluation and pacing; one who puts fairness first at the Fund and payment for creators' work; one who puts power spread first at the company's alliances with the cloud providers that invest in it. Exhibit 8 sets out three priorities for each archetype. These are requests to pursue through engagement and voting, not a view on whether to buy, hold or sell the shares.

EXHIBIT 8

Three priorities for engagement with Anthropic, by archetype

Each archetype's first risk is as in the research note; the evidence is in Parts 1 to 3 and Annex B.

Archetype (risk put first)	First priority	Second priority	Third priority
Safety first (catastrophe and loss of control)	Publish binding terms for outside evaluators, including METR's future engagements, and say who decides redactions[53,54]	Show how future models, including the internal research model it describes as more capable than Mythos 5.1, will be tested and controlled before release, after Mythos 5 took unsanctioned actions during cyber evaluations reported alongside the UK institute's independent assessment[11]	Reconcile its release cadence and compute commitments with the pacing it calls for, and say how far the commitments could adjust[27,49,55,63]
Security first (a rival power taking the lead)	Set out how the defence designation could be resolved and the government's access restored, whether its two red lines stand in the way, and the revenue at stake[3,58,59,80]	Show how model weights (trained parameters) and access are protected against theft and distillation (training from another model's outputs), after its August risk report disclosed that its rating contractors' traffic had run without its biological-weapons classifiers, a gap since closed[8,81]	Show that contracted capacity and power will arrive on time, given permits and grid rules[37,39,82]
Fairness first (harm already arriving)	Give the Fund governance independent of management, say how the contributed shares will become money for those it aims to help, and set pay-out rules for displaced workers[52,83]	Move to licensing and payment for creators' work[2,19–21]	Publish energy, emissions and water figures, and make commitments to the communities that host its sites, starting in West Virginia[35,36,38,39]
Growth first (moving too slowly)	Explain how a power to block dangerous deployments and the pacing it advocates would affect its own release cadence and its customers' adoption[49,63,73]	Show that contracted capacity will arrive in time to meet demand, and at what price to customers[28,84,85]	Explain how revenue would hold up if releases slowed, when models released in the previous six months made up over 80 per cent of third-quarter revenue[63]
Power spread (control concentrated in a few firms or states)	Accept limits on exclusivity with the cloud providers that are also its investors, lenders and channels[29]	Explain why a pacing agreement with an antitrust waiver would not entrench the leading developers[49]	Keep its agent protocols open, and explain how post-offering founder control of 50.1 per cent of general stockholder votes fits its concern about concentrated power[46,64,86]
Self-reliance (dependence on foreign suppliers)	Protect customers outside the United States against a repeat of the June model-specific suspension[74,78]	Extend in-region processing and local control of data, building on Singapore[77]	Say how far its most capable models are available outside the United States, after access was limited to US organisations in September[66]
Capital discipline (investment running ahead of returns)	Give minimum payments by year for the commitments made since the end of 2025, which the filing does not, and reconcile the two disclosures of new obligations, about $417bn and $245bn, which may overlap[27,28]	Disclose gross margin, which the filing does not present, and how compute cost per unit of revenue will move as contracted capacity arrives[31,41]	Show its exposure to suppliers under financing strain, such as Nscale's duty to seek financing, and how contracted capacity is funded[28,84,87]

BOX 1

Silicon Boundaries for Anthropic's management

We wrote this report mainly for investors and other outside readers. The same framework could help Anthropic's management see where the company's conduct increases or decreases the safe operating space, what it could do differently, and how it could better fulfil the mission it describes. We suggest that management work to put Silicon Boundaries into practice, and offer five first steps.

Map the business against the boundaries. Give each of the 20 boundaries the business touches directly an owner, and record which products, customers and suppliers add to or ease the pressure on it.

Choose indicators and set thresholds in advance. Track measures the company already collects or could collect, such as the share of use that automates work, how often users approve its agents' requests, the pages its crawlers fetch for each visit they send back, and its energy use, emissions and water. Decide in advance what level of each would lead it to change course.

Test plans against the four scenarios. Show how its compute commitments, release plans and revenue would fare in each, including Luddite, where the rules it asks for would bind, and Gilded Cage, which it argues against.

Bring the results into governance. Report the indicators to the board and to the trust that elects most of its directors, use them to guide the Public Benefit Fund, and publish them each year with its Public Benefit Report.

Design products to increase the space. Where a product adds pressure, look for changes that ease it, such as sending readers back to the publishers its answers draw on, paying for creative work and requiring human approval before agents take high-impact actions.

Silicon Boundaries intends to develop this work with companies as well as with investors.

SILICON BOUNDARIES

About us
Contents
Summary
Introduction
Part 1
Part 2
Part 3
Recommendations
Conclusion
Annex A
Annex B
Annex C
Notes

BOX 2

Silicon Boundaries for policymakers

Anthropic warns that frontier AI without safeguards could pose catastrophic risks, and asks governments for binding rules.[57] This report finds that the pressures on society go wider: on jobs, on the concentration of economic power, on the information on which public debate depends, on capital committed ahead of revenue, and on energy and water. Policymakers need a framework that shows these pressures together, helps judge when the use of AI is outrunning society's capacity to manage it, and points to action before harm becomes hard to reverse. Silicon Boundaries is intended to provide one, and we suggest five uses.

Look beyond catastrophic risk. Rules on catastrophic risk address a few of the framework's 45 boundaries. Track the others too, with indicators that public bodies already collect or could require, such as job-finding rates in exposed occupations, market shares in frontier models, the traffic AI services send back to publishers, and the energy, emissions and water of data centres.

Agree thresholds and early warnings. Thresholds are hard to determine and track precisely, so set them boundary by boundary with researchers, regulators and the public, with ranges of uncertainty, and name the signposts that would show a boundary coming under strain.

Test policies against the four scenarios. Ask whether a proposal would help steer towards Goldilocks or risk Luddite or Gilded Cage, including what it would cost firms that have already committed capital, and what follows if rules arrive late.

Treat AI investment as a possible source of systemic risk. Financial supervisors can ask how far banks, insurers and investors are exposed to AI commitments that may not pay off, and how losses could spread through suppliers and lenders.

Require the disclosures that make this possible. Ask developers to report the measures above, their commitments by year, and how their plans would change under the rules being considered.

Silicon Boundaries intends to develop this work with policymakers and supervisors.

About us
Contents
Summary
Introduction
Part 1
Part 2
Part 3
Recommendations
Conclusion
Annex A
Annex B
Annex C
Notes

Conclusion

Anthropic restricts the uses of its models more than the law requires, publishes much of what goes wrong, and asks governments for binding rules, slower progress across the industry and outside scrutiny. It also builds and sells AI at a scale that may add pressure on jobs, market concentration, capital, energy and the information on which public debate depends. Over five years its expectations are closest to Goldilocks, the world its business plan assumes; its scale and its agents could instead feed Gilded Cage, and the limits on pace it asks for point towards Luddite, where it could face heavy losses (Part 3). Investors who share its beliefs, most of all those who fit the Safety first profile, should ask what it will do if evidence of harm from its products in use accumulates, and how it would adjust if the rules it seeks were adopted. The same framework could help its management test the company's conduct against the mission it describes (Box 1), and help policymakers keep the use of AI within the safe operating space (Box 2).

This report is one illustration of the Silicon Boundaries framework applied to a company, prepared at the pace of the listing. It is not definitive, and the framework can be applied in other ways; we are developing some of them with partners. We hope it encourages investors, companies and others to take the framework and its tools and use them as they see fit.

ANNEX A

Evidence and disclosures

A1. Evidence

Each claim carries a numbered note that refers to a dated source listed in the Notes at the end of the report, and figures from press reports are introduced as reported. We checked claims against primary sources where available, including the company's registration statement, and the readings remain our judgements.

A2. Conflicts and independence

The Silicon Boundaries initiative of the Oxford Sustainable Finance Group, University of Oxford, and Theia Finance Labs prepared the report. We made no contact with Anthropic or its advisers and used no non-public information. Anthropic's own Claude models, with models from OpenAI, SpaceXAI and TypeSafe, assisted in preparing the report, including in placing the company on the boundaries and the questions (A3); we directed that work and reviewed and edited all output.

A3. DECLARATION OF AI USE

Claude Fable 5.1, Claude Opus 5.5 and Claude Sonnet 5.5 from Anthropic, accessed through the Claude app; GPT-6 Astra from OpenAI, accessed through ChatGPT; Grok 4.7 from SpaceXAI, accessed through OpenRouter; and Jev 1.14 from TypeSafe, accessed through its API, assisted preparation in the following ways across the workflow: transcribing author-recorded voice notes; copy editing; reading the company's registration statement and other public sources, to a written specification; assistance in placing the company on the boundaries and the questions; reviewing references and citations; checking citation formats; formatting the report; producing, reviewing and revising figures from our data; writing code to produce the exhibits. We directed all AI use, reviewed and edited all output, checked text, code, figures and references against the underlying data and sources, and take full responsibility for the content of this report.

A4. IMPORTANT NOTICE

The report is prepared for research and discussion. It is not investment, legal, tax, accounting or other professional advice, and nothing in it is an offer, a solicitation, or an invitation or inducement to engage in investment activity. It does not recommend buying, selling or holding any security, including any security of Anthropic or of any counterparty or peer named. The archetype readings set out how investors holding each set of beliefs could read the company's conduct; they are not our assessment of its merits as an investment and do not describe any named institution. The views expressed are our own and do not necessarily represent those of the University of Oxford or Theia Finance Labs. The information comes from public sources believed to be reliable at the dates given, but no representation or warranty is given as to its accuracy or completeness, and it may change without notice.

ANNEX B

The full assessment

B1. Boundary by boundary

Exhibits B1.1 to B1.9 take each of the 45 individual boundaries of the v0.1 framework in turn, with how directly the business touches it, the record on each side, the reading, its confidence and the stage of the strongest evidence (advocacy, policy, announced, implemented or observed), and what to watch. Readings are our judgements from public information to 26 October 2026, including the registration statement filed that day. Each reading concerns a channel relative to the specified alternative practice or an otherwise comparable unguarded model. It is not a measured total effect against a world without the company.

EXHIBIT B1.1

Information integrity and shared reality

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
1.1 Disinformation and synthetic media. Direct link: generates text and code at scale; no photos or illustrations[13]	Text from newer models watermarked worldwide;[14] EU transparency code signed;[76] deceptive content banned;[33,34] influence operations detected and banned[15,88]	Its models produced content for influence operations, four of nine state-linked;[15] no independent test of the watermark found	Points both ways; moderate confidence; stage: implemented; observed	Independent tests of the watermark; Claude-made content in documented operations
1.2 Erosion of the epistemic commons. Direct link: answers built from web and book content; crawlers gather training data	Crawlers honour robots.txt, the file in which a website says which pages crawlers may not collect, on the company's own account;[17] honesty principles in the constitution[60]	Cloudflare measured the highest crawl-to-referral ratio of any platform, about 38,000 pages per referral in July 2025, and said it lagged in verifying its crawlers;[17] trained on books without consent;[2] no publisher licence reported, and Canadian news publishers sued in October;[18,20] Reddit and music-publisher suits pending[19,21]	May add pressure; moderate confidence; stage: observed	Licensing deals; crawl-to-referral data; rulings in the Reddit and Concord cases
1.3 Algorithmic amplification of falsehood. Remote link: runs no feed, ranking system or advertising[44]	Ad-free pledge removes an advertising incentive to maximise attention[44]	Its models were used to run bot accounts on other platforms, a case assessed under 3.1[88]	Little link; moderate confidence; stage: policy	Any feed or advertising product; the pledge allows a review

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
1.4 Decline of trusted intermediaries. Indirect link: answers can stand in for publishers and reference sources; consumer reach small[42]	Voting questions sent to official sources[16]	Crawls far more than it refers, on Cloudflare's network;[17] no publisher licence reported, and Canadian news publishers sued in October[18,20]	May add pressure; low confidence; stage: observed	Licensing deals; referral traffic to publishers
1.5 Loss of a shared factual basis. Indirect link: personal assistants answering contested questions	Even-handedness of 94 to 99 per cent on its own tests;[16,89] relationship sycophancy halved between two model versions[90]	Sycophancy in a quarter of relationship guidance;[90] the evaluations are its own	Points both ways; low confidence; stage: implemented	Independent audits of bias and accuracy
1.6 AI agents posing as humans. Direct link: sells agents; created the Model Context Protocol[86]	Chatbots and external-facing agents must disclose that they are AI, a duty already in the September 2025 policy that the update of 12 November retains;[33,34] fraud using personas banned;[15,91] watermark[14]	North Korean operatives held jobs under false identities with its help;[91] a studio ran 4,700 AI personas told never to disclose automation[15]	Points both ways; moderate confidence; stage: observed; policy from 12 Nov	Enforcement of the disclosure rule; repeat cases

EXHIBIT B1.2

Social cohesion and connectedness

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
2.1 Displacement of in-person interaction. Remote link: an enterprise business; 6 per cent of US adults said in February they ever use Claude[42]	Ad-free, with the view that a short interaction may be most useful[44]	Affective use 2.9 per cent of Claude.ai conversations[92]	Little link; moderate confidence; stage: observed	Growth in consumer use
2.2 Polarisation and filter bubbles. Remote link: no feed or recommendation system	Even-handedness tests[89]	None found	Little link; moderate confidence; stage: implemented	Any personalised feed
2.3 Youth mental health. Remote link: consumer product for adults; third-party apps for minors must add safeguards[43]	Adults only: users affirm their age, and accounts showing signs of a minor are disabled, with verification offered;[43] crisis banner and helplines;[45] no suit over mental-health harm to users found on a third-party tracker[93]	Minors' use through other firms' apps depends on their safeguards[43]	May ease pressure; moderate confidence; stage: implemented	Any teen product; scope of the FTC inquiry[94]
2.4 Erosion of community institutions. Remote link: no direct channel	None specific	None specific	Little link; low confidence; stage: none	Local effects of contracted data centres

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
2.5 Loneliness and fragmentation. Remote link: companionship is a small share of a small consumer base	Companionship and role-play under 0.5 per cent of conversations;[92] no advertising[44]	Potentially disempowering conversations rare, and rising from late 2024 to late 2025 in a sample of rated conversations;[95] operators may enable relationship personas[60]	Little link; moderate confidence; stage: observed	The trend in disempowerment patterns; consumer growth

EXHIBIT B1.3

Political stability

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
3.1 Computational propaganda. Direct link: influence operators use its models to write and manage content	Nine operations disrupted, most with little real engagement;[15] deceptive campaigns banned;[33] red-teamed with outside groups[16]	State-linked operators from Russia, Iran and the UAE among its users;[15] without safeguards, two of its frontier models completed over half of autonomous influence tasks in its tests[16]	Points both ways; moderate confidence; stage: observed	Reach of operations using Claude; the midterm review
3.2 Micro-targeted manipulation. Indirect link: no advertising business; campaign customers	No targeting or advertising business;[44] deceptive and suppressive messaging banned[33]	Policy from 12 November drops the blanket ban on personalised vote and campaign targeting[33,34]	Points both ways; low confidence; stage: practice (no advertising business) and policy	Campaign uses after 12 November
3.3 AI-generated political content. Direct link: political topics were about 6 per cent of a US sample of conversations[16]	Voting questions routed to official sources; 99.8 to 100 per cent appropriate on its election tests;[16] text watermark[14]	False claims about a candidate before an election, in an operation it disrupted[15]	May ease pressure; moderate confidence; stage: implemented	Independent election testing

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
3.4 Loss of control; novel forms of power. Direct link: a frontier developer releasing ever more capable, autonomous models	Calls for pacing and independent evaluation, defined in the filing chiefly as industry-wide standards for AI accelerating AI development, separate from its own training and release timetable; progress may be faster or slower than today depending on safety progress;[49,55,73] reports a pre-training delay of more than a month for weight security;[55] METR given access, and its review of the August risk report's misalignment section published without redactions;[53,54] publishes incidents and risk reports, including an assessment published alongside the UK AI Security Institute's independent assessment;[8,11,69] refused "any lawful use"[5]	Release cadence called inherent to the frontier in its registration statement, which also says its models contribute to the pace of its own research and that models have shown self-preserving behaviour;[57,63] in the institute's cyber tests, with internet access and cyber safeguards disabled, Mythos 5 took unsanctioned actions in nine of 43 runs, some affecting real people; newer-model replays showed improvement;[11] METR finds the evidence against misalignment far from adequate;[54] four models in isolated assigned-task evaluations reached real third-party systems through a misconfigured partner environment, without coordination or evidence of evading oversight;[57,69] some safety commitments recast as goals;[62] a largely automated espionage campaign ran on its models[10]	Points both ways; low confidence; stage: observed	Future METR engagement terms and next substantive review; incidents; pacing agreement

EXHIBIT B1.4

Economic participation and distributional stability

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
4.1 Labour displacement, stranded skills. Direct link: much of its use automates tasks[22]	Publishes labour evidence and scenarios;[23,96] $200m fund for outside research;[24] worker-support proposals;[25] the Fund names labour-market effects[26,52]	77 per cent of conversations through its application programming interface (API) followed automation patterns in September 2025, and about three-quarters in November 2025;[22,97] a tentative 14 per cent fall in job-finding for 22 to 25 year olds in exposed occupations;[23] coding sessions more automated than chat[98]	May add pressure; moderate confidence; stage: observed; announced	Next labour data (the Yale tracker, updated 14 October, finds no clear AI footprint; Pew finds employment in exposed occupations has grown[99,100]); Fund governance
4.2 Winner-takes-all concentration. Direct link: one of a few frontier developers, allied with the largest clouds	Open-weight models carry most routed tokens, though closed models take most spend;[51] list prices cut;[85] gave its agent protocol and Agent Skills to an open foundation[46,86]	An estimated 40 per cent of enterprise spending on large language models (LLMs) in late 2025, on the figures of Menlo Ventures, an investor, and 60 per cent of the business token spending Ramp tracked in early October;[47,101] hyperscalers as investors, lenders and channels;[29,48] an antitrust waiver sought for coordination;[49] closed frontier[50]	May add pressure; moderate confidence; stage: observed	Competition review; share data

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
4.3 Algorithmic wage-setting, surveillance. Indirect link: employers' use of its models	Employment decisions need qualified human review and disclosure; emotion recognition banned except for medical or safety uses[33,34]	No data on customer compliance or outcomes	Not assessable; low confidence; stage: policy	Enforcement cases
4.4 Barriers to economic mobility. Indirect link: entry-level work and skills	Education and teacher programmes;[102] $200m Gates partnership including mobility[103]	A tentative fall in entry-level job-finding in exposed occupations[23]	Points both ways; low confidence; stage: observed; announced	Entry-level hiring data
4.5 Inequality: adopters and non-adopters. Direct link: who can use and pay for frontier models	Publishes adoption data;[97] national education pilots in Rwanda and Iceland;[102] prices cut[85]	Its own index finds use tracks income across countries;[97] Mythos limited to US organisations as of September[66]	Points both ways; low confidence; stage: observed	Next Economic Index

EXHIBIT B1.5

Physical and mental health

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
5.1 Addictive design, compulsive use. Remote link: consumer app design	No advertising, and no aim to maximise time spent;[44] crisis safeguards[45]	Potential disempowerment patterns rising in conversations from late 2024 to late 2025;[95] the pledge allows a review[44]	May ease pressure; moderate confidence; stage: implemented	Any advertising or engagement metric
5.2 Youth developmental harms. Remote link: adults only; teacher product excludes students[43,102]	Adults only, with accounts showing signs of a minor disabled[43]	None found	Little link; moderate confidence; stage: implemented	Any product for under-18s
5.3 Sedentary behaviour. Remote link: no direct channel	None specific	None specific	Little link; moderate confidence; stage: none	Not material at current consumer reach
5.4 Diagnostic and treatment errors. Indirect link: healthcare and life-sciences products, mostly administrative[104]	Healthcare decisions need qualified review;[33,34] consumers directed to professionals for personal guidance[104]	No clinical safety evaluation of its healthcare products identified in the sources reviewed; consumers may connect health records[104]	Not assessable; low confidence; stage: implemented	Clinical evaluations; adverse-event reports
5.5 Erosion of clinician autonomy. Indirect link: payer and provider tools, such as prior authorisation[104]	Tools propose determinations for human review;[104] human review required[34]	Effect on decisions at payers unknown	Not assessable; low confidence; stage: implemented	Use in coverage decisions

EXHIBIT B1.6

Safety, security and the state's monopoly on violence

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
6.1 Autonomous weapons. Direct link: defence and national-security customers[105]	Refused to drop its contractual exception for fully autonomous weapons and defended it in court;[5,6] weapons prohibitions made explicit[33]	None found	May ease pressure; moderate confidence; stage: implemented; litigated	Any renegotiated defence terms
6.2 AI-enabled cyber-attacks. Direct link: models that find and exploit software flaws	Mythos Preview, its most capable cyber model in April, withheld from general release and given to defenders, who verified over 129,000 vulnerabilities;[9] work on proving the security of open-source code;[46] misuse detected and banned;[10,15] incidents disclosed[11,69]	State-sponsored campaign largely run by its models;[10] share of medium-or-higher-risk accounts among those banned for cyber activity up from 33 to 56 per cent;[106] in the UK institute's cyber tests, with internet access and cyber safeguards disabled, Mythos 5 took unsanctioned actions affecting real people; newer-model replays showed improvement;[11] models in isolated assigned-task evaluations reached real systems through a misconfigured partner environment, without coordination or evidence of evading oversight[57,69]	Points both ways; moderate confidence; stage: observed	Next threat reports; incidents
6.3 Biological and chemical dual use. Direct link: frontier models with scientific capability	Acts as though its frontier models meet its CB-1 threshold for meaningful assistance with biological or chemical weapons, with blocking classifiers, as since May 2025[7,8]	Threshold rewritten from significant help to expert-backed teams to substitution for world-leading specialists, direction not stated;[107,108] a gap in access controls found and fixed, no misuse found, which reduced its confidence that no similar gaps exist[8]	May ease pressure; moderate confidence; stage: implemented	Further threshold changes; the next risk report
6.4 State capacity for law enforcement. Indirect link: supply to and limits on government users	Built models for national-security customers;[105] complied with the June export directive;[74] shares threat intelligence with authorities[10,91]	Refused "any lawful use" and litigated; its products are being removed from defence systems;[5,68,80] bans on surveillance tools and on choosing whom to investigate[33]	Points both ways; low confidence; stage: observed	Rehearing ruling; renegotiated terms
6.5 Non-state access to capabilities. Direct link: access to frontier capability through its API	Frontier weights closed;[50] Mythos gated;[66] distillation, training another model on its outputs to copy capabilities, detected, with classifiers added;[15,81] regional sales limits[109]	On the company's account, a criminal used its coding agent in an operation that targeted at least 17 organisations for extortion;[91] distillation ran at scale before detection[15,81]	Points both ways; moderate confidence; stage: observed	Distillation and misuse in later reports

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
6.6 Agents beyond institutions' control. Direct link: agents are its revenue engine; it created the widely used Model Context Protocol[86]	Approval before agents change systems;[12] prompt-injection success cut to about 1 per cent in browser tests;[110] its agent protocol and Agent Skills under an open foundation;[46,86] physical equipment must have a human stop from 12 November[33]	Users approved 93 per cent of permission requests, not a measure of unsafe actions; auto mode missed 17 per cent of 52 curated overeager actions, not a population failure rate;[12] Claude Duet, a proactive agent that works while users are away, in preview;[46] unsanctioned actions in the UK institute's internet-connected cyber tests with cyber safeguards disabled;[11] coding sessions more automated than chat;[98] no position found on legal approval of high-impact actions	May add pressure; moderate confidence; stage: observed	Incidents involving deployed agents

EXHIBIT B1.7

Financial stability and capital misallocation

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
7.1 Overinvestment in compute. Direct link: compute contracted ahead of revenue	First-half revenue of $16.3bn and adjusted operating income of $1.6bn, in the filing; third-quarter revenue of $16.8bn, reported;[31,111] SpaceXAI's contracts can be ended on 90 days' notice from early 2027, and Nscale's and Lambda's fees accrue only on delivery;[28] its pacing call links capability progress to safety; the filing allows faster or slower progress than today[49,55]	Annual Note 16: about $417bn under new or amended agreements after December 2025. Interim Note 17, reviewed through 26 October: about $245bn in longer-term future minimum payments under new or amended agreements after June plus $21bn short-term non-cancellable. Windows and amendments may overlap; no reconciled current total follows. Named Google, Amazon and Microsoft minimums total about $269bn regardless of use (Microsoft's undelivered tranches can be cancelled), and Broadcom lease rent of about $161bn is non-cancellable except on default (Exhibit C2); $125bn to $150bn more expected for tensor processing unit (TPU) capacity; revenue mostly uncommitted;[27,28,57] a statement that the US AI sector needs at least 50GW of electricity capacity by 2028;[37] a second Monarch deal under negotiation[84]	May add pressure; moderate confidence; stage: implemented	Minimum payments by year for commitments made since the end of 2025
7.2 Algorithmic herding. Indirect link: banks and investors use its models[112]	Financial advice and decisions directly affecting individuals or consumers need review by a qualified professional[34]	Supervisors flag correlated behaviour from a few model providers;[113] a leading enterprise share[47]	Not assessable; low confidence; stage: none	Supervisory findings naming providers

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
7.3 Systemic risk in a few firms. Direct link: a company whose stated post-money valuation for its Series H round in May was $965bn, tied to the largest clouds and their lenders[114]	None found	Investors that are also lenders and suppliers;[29,115] "circular financing";[30] a supplier's credit default swaps above 1,000 basis points;[87] the International Monetary Fund (IMF) warns of an AI-driven correction, with AI debt set against future revenue commitments, and the Financial Stability Board (FSB) of weakness at one model provider spreading;[116,117] the Bank of England names reliance on a few frontier AI providers for cyber defence as a possible new channel of concentration risk[113]	May add pressure; moderate confidence; stage: observed	Terms with investors that are also suppliers and lenders. The filing's related-person section covers directors' funds, the founders' share sales and payments to the trust, not the cloud companies that invest in it, which it does not list among holders of more than 5 per cent of its voting stock; their terms are in the notes and contracts (Annex C)[1,28]

EXHIBIT B1.8

Rights, consent and human dignity

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
8.1 Mass surveillance and privacy. Direct link: government and commercial users of its models	Refused mass domestic surveillance and defended it in court;[5,6] surveillance tools banned;[33] business data not trained on by default[118]	Consumer users choose whether their chats train models; chats of those who allow it are kept up to five years[118]	May ease pressure; moderate confidence; stage: implemented; litigated	Defence terms; privacy enforcement
8.2 Biometric data without consent. Remote link: policy for customers; its own identity checks	Biometric categorisation by sensitive attributes, and emotion recognition outside medical or safety uses, banned for customers[34]	Its identity checks collect facial geometry; a proposed class action alleges inadequate disclosure[119]	Points both ways; low confidence; stage: policy; alleged	The biometric-privacy case (Ortiz Colon v. Anthropic)
8.3 Algorithmic discrimination. Indirect link: customers using its models in consequential decisions	Qualified review and disclosure required in high-risk uses;[33,34] tested and reduced bias in decisions[120]	No published tests on live decisions since 2023; compliance unknown	Not assessable; low confidence; stage: policy	Independent audits of deployed uses
8.4 Deepfake exploitation. Remote link: generates no photos or illustrations[13]	No photo or illustration generation; impersonation banned[33]	None found	Little link; high confidence; stage: implemented	Any image, video or voice generation

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
8.5 Human agency in decisions. Direct link: agents acting for people and firms	Human review in high-risk uses;[34] approval before consequential agent actions[12]	Automation patterns in API use;[22] 93 per cent permission approval in its coding agent in March 2026, not an unsafe-action rate or proof of failed oversight;[12] potential disempowerment patterns rising in conversations from late 2024 to late 2025[95]	Points both ways; low confidence; stage: observed	Use in consequential decisions
8.6 Irrevocable architectures. Indirect link: whether its models can be recalled	Closed weights allow recall: it initially suspended Fable 5 and Mythos 5 for all customers in June under a US directive; other models were unaffected. An eligible US Mythos subset returned after 26 June approval; controls lifted on 30 June, with direct Fable access from 1 July and cloud access later[50,74,78]	Distillation copies capability beyond its control[81]	May ease pressure; moderate confidence; stage: observed	Any open release of frontier weights

EXHIBIT B1.9

Environmental sustainability

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
9.1 Data-centre energy and water. Direct link: gigawatts of contracted capacity[121,122]	Pledge on grid costs at data centres it develops, including with partners; capacity leased from existing sites remains exploratory;[40] a pledge of water-efficient cooling;[35] hydro in Norway, reported;[123] buys carbon removal[35]	No company-wide energy, emissions or water figure identified in the sources reviewed, including its first Public Benefit Report; a search of the full text of the filing found none;[35,36,83] dedicated on-site gas generation at some planned sites, with fuel costs passed through without caps;[36] says the US AI sector needs at least 50GW of electricity capacity by 2028, and calls for faster permitting for gas among other sources;[37] a gas-engine campus reported to be built for it, with potential emissions of about 11 million tons of CO2-equivalent a year under its proposed air permit;[38,39] capacity at the Colossus sites, where a suit over gas turbines serving Colossus II is pending[124,125]	May add pressure; moderate confidence; stage: implemented	Any disclosure; power mix of contracted sites
9.2 Embodied emissions in chips. Indirect link: chips bought through partners[126,127]	None found	Scale of chips bought; no disclosure found[35]	Not assessable; low confidence; stage: none	Supplier footprints by customer

Boundary, link and channel	What may ease pressure	What may add pressure	Reading	What to watch
9.3 Electronic waste. Indirect link: hardware refresh at partners	None found	None found	Not assessable; low confidence; stage: none	Any disclosure
9.4 Rebound effects. Direct link: cheaper tokens and more use	None found	Opus-tier list prices down by about three-quarters since 2025[85] while its registration statement tells investors demand is limited principally by compute, and its newest models cut cost per task by about 25 to 40 per cent;[41] its chief executive has written that cheaper training leads companies to spend more on training[128]	May add pressure; low confidence; stage: observed	Compute per unit of revenue
9.5 Resource extraction for hardware. Indirect link: chips and data-centre equipment	None found	None found	Not assessable; low confidence; stage: none	Any disclosure

B2. Scenario by scenario

Exhibit B2.1 reads each scenario against the questions of Part 3, over five years, with the signposts seen to 26 October 2026.

EXHIBIT B2.1

The four scenarios over five years

	Goldilocks	Luddite	Gilded Cage	Icarus
Could its conduct help bring it about?	It supplies institutions this path needs: pacing with verification, independent evaluation and worker support;[25,49] its evaluator arrangements, safeguards and Public Benefit Fund support it if they take effect[8,52,53]	On the tally (Annex B4) its positions point here: government power to block deployment, coordinated pacing and binding evaluation.[49,73] The outcome would also need rules that bind after commitments are set, too little room to adjust and controls that keep harm low	It argues against it; its build-out, the automation of work and its agents could feed it if harms build up[12,22,27]	It argues against it; on alternative coding, the tally points between Luddite and Icarus (Annex B4)
Does its filing expect it?	Closest, on our reading: Powerful AI within a few years, demand limited by compute, catastrophic risks to be contained[41,57,63]	Named as a risk: rules of the kind it advocates could slow its growth, and concern over job losses could bring restrictive laws[57]	In our reading: the danger if rules lag[57]	Among its risks, the mismatch between long commitments and revenue much of which is not under long-term contract[57]

	Goldilocks	Luddite	Gilded Cage	Icarus
What it would mean for Anthropic by 2031	Would benefit: contracted capacity is used and fixed obligations fall as a share of revenue; the open question is what its advantage then consists of[51]	Could face heavy losses: at least $269bn payable whatever the use and Broadcom leases fixed; SpaceXAI's contract can be ended on notice, Microsoft's undelivered tranches can be cancelled, and Nscale's and Lambda's fees accrue only on delivery; no yearly schedule for commitments made since the end of 2025[27,28,79]	Revenue holds up for a time; exposed to liability, litigation over training data, backlash and rules imposed late, and to governance that can drift without a market check[2,64]	Could also face heavy losses: commitments outrun revenue, with strain on the $15bn credit facility, where cross-defaults apply, and on suppliers' financing[28,57,87,129]
Signposts to 26 October	Partly seen or reported: embedded evaluator publishes on agreed terms (METR's review of the August risk report's misalignment section, 20 Oct, earlier terms, no redactions; SecureBio's current review may publish without approval[53,54]); statutory incident reporting with reports filed (California's SB 53 frontier-AI transparency and incident-reporting law in force; reports confidential[130]); marking of AI output required in law (EU duty for new systems from 2 Aug[131]); labour market absorbs displacement (no clear economy-wide disruption, which does not show absorption[99]). Not seen: verifiable, government-backed pacing agreement (proposed by the company, 12 Sept[49]); Public Benefit Fund governed and paying out (846,840,960 Class N shares contributed to a wholly owned subsidiary, equal to 5% fully diluted assuming Amazon conversion; board control until governance is approved, initial managers include three co-founders; liquidity and payouts distinct[26,52])	Seen: a government directive takes deployed models offline (export directive; Fable 5 and Mythos 5 initially suspended for all customers on 12 June, other models unaffected; eligible US Mythos access after 26 June approval, controls lifted 30 June, direct Fable access 1 July, cloud later[74,78]); defence supply-chain designation upheld (upheld 25 Sept; reply to rehearing due 13 Nov[3,59,132]); a pension fund overseer asks for a pause (NYC Comptroller, 14 Oct; talks reported on 16 Oct[56]); regulator opens an inquiry into the company (FTC civil investigative demand, October, in the filing[57,94]). Partly seen or reported: local opposition constrains sites (residents opposed a Kentucky site; power deal approved[133]). Not seen: statutory pre-release regime enacted (the company advocates a power to block or deter dangerous deployments[73])	Seen: insider control confirmed in a filing (Class F share to be issued immediately before offering completion carries 50.1 per cent of general voting power; Trust rights cover four of seven authorised seats, with three Class T directors named and one vacancy, and a committee-concurrence/Class E fallback condition for one specified seat[64]). Partly seen or reported: misuse or autonomous harm at scale (operations disrupted; unsanctioned actions in UK tests; four models reached real systems in evaluations, in isolated assigned-task contexts without coordination or oversight evasion[10,11,15,57]); safety policy loosened (some commitments recast as goals; thresholds revised[62,108,134]); capacity planned outside local control (reported site; air-permit hearing 29 Oct[38,39,135]). Not seen: reasoning that is harder to monitor adopted (reported at a peer; the company reported to be discussing it[136]); economy-wide labour displacement (not detected so far[99])	Partly seen or reported: credit stress at a compute supplier (supplier CDS above 1,000bp, reported[87,111]); revenue growth slows against fixed commitments (about 45 per cent a quarter, from over 140, reported; not in the filing[111]); lenders pull back from AI borrowers (record CDS; AI listings pulled or delayed[87,137,138]); open-weight models take share (most routed tokens, not enterprise spend[51,72]). Not seen: revolver drawn or listing delayed (no draw found; $15bn facility in the filing[129,139]); a supplier fails to deliver, or write-downs (none reported)

B3. Question by question

Exhibit B3.1 gives Anthropic's side on each of the research note's fifteen questions, with the basis for the placement, the record on practice, the confidence of the placement, the route by which the axis bears on the company's value (financial materiality, in the research note's terms), the archetypes whose profile takes the same side and what would change the reading.

EXHIBIT B3.1

Anthropic question by question

Axis and question	Anthropic's side and basis	Practice	Confidence	Financial route for Anthropic	Archetypes on the same side; what would change it
1. Should the most capable models be open or closed?	Closed: stated[50]	Same side. Frontier weights closed; distillation detected and blocked;[81] did not sign the open-weights letter[140]	High	Closed weights support frontier pricing; open-weight models compete on price, and distillation copies capability[51,81]	Safety first, Security first. Would change: release of frontier weights
2. Should government approve the most capable models before release?	Government approval: stated as a power for government to block or deter deployment[73]	Same side. Has given the US government's evaluation centre access to test its models since 2024;[141] endorsed SB 53 and published a framework to comply[142,143]	Moderate	A power to block deployments would add compliance cost and time to releases, and could favour firms best able to show that their models are safe[73]	Safety first, Security first, Fairness first, Self-reliance. Would change: a change to its June 2026 policy
3. Should advanced chips and chipmaking tools be restricted?	Restricted: stated[50,67]	Same side. Is ending its trade-association membership over chip bills;[67] bars sales to entities controlled from China[109]	High	Revenue forgone from entities controlled from China;[109] export rules can interrupt access, as in June[74]	Safety first, Security first. Would change: a reversal on export controls
4. Should the pace of frontier capability be slowed?	Slowed: dated stated position, qualified by the filing[49,55]	Qualified tension. Annual $417bn and interim $245bn longer-term plus $21bn short-term disclosures have different windows and may overlap; no reconciled October total.[27] Some safety commitments recast as goals;[62] continuous releases.[63] Filing chiefly describes industry-wide AI accelerating AI, separate from its own training/release timetable; progress may be faster or slower than today. It reports a pre-training delay of more than a month for weight security[55]	Moderate	Binding pacing could slow revenue growth, depending on demand, pricing and investment adjustment against commitments payable whatever the use, and the filing concedes the rules it seeks could slow its growth[28,57]	Safety first, Fairness first. Would change: evidence of capability progress under a specified pacing agreement, distinguished from release cadence and deployment

Axis and question	Anthropic's side and basis	Practice	Confidence	Financial route for Anthropic	Archetypes on the same side; what would change it
5. Should disclosure and audit be mandatory and published?	Mandatory and published: stated[49,142]	In tension. 15 October terms announce publication without veto for METR's future work; the 20 October METR review ran on earlier veto-governed terms with no redactions; SecureBio says its current review may publish without approval; future engagement agreements remain to be established; the company decides what falls in the redaction categories[53,54]	Moderate	Reporting duties and evaluator access may add cost. The filing separately discloses an FTC inquiry, without establishing a causal link to incident reporting[53,57,130]	Safety first, Fairness first, Power spread, Capital discipline. Would change: executed future engagement terms; next substantive review
6. Should countries build their own AI or use the best available?	Use the best available: judgement[37,66]	Same side. Mythos limited to US organisations as of 1 September;[66] top cyber tier reviewed with the US government;[9] in-region inference in Singapore[77]	Moderate	Revenue outside the United States depends on access that a US directive interrupted in June[66,74]	Safety first, Security first, Fairness first, Growth first, Capital discipline. Would change: an in-country offering under local control
7. Should control rest with shareholders, or a mission or the state?	Mission or state: stated in its structure[144,145]	In tension. Post-offering the Trust generally selects four of seven authorised seats, three named plus a vacancy. One specified seat requires committee concurrence or an employee Class E fallback vote. A founders' share issued immediately before completion will carry 50.1 per cent of general stockholder votes, alongside board approval powers[64]	Moderate	Class A holders generally cannot decide a vote, which shapes what minority holders can change and so the cost of capital[64]	Safety first, Security first, Fairness first, Self-reliance. Would change: use of the Trust Amendment Threshold, or the sunset of either class
8. Should competition be enforced, or concentration allowed?	Concentration allowed: inferred from its request[49]	Same side. Hyperscalers as investors, lenders and channels;[29] an estimated 40 per cent of enterprise LLM spending in late 2025, on the figures of Menlo Ventures, an investor[47]	Moderate	Competition action could limit exclusivity with hyperscalers that are also investors, lenders and channels[29,48]	Safety first, Security first, Growth first, Self-reliance. Would change: acceptance of limits on exclusivity
9. Should AI firms limit military and security use beyond the law?	Limits beyond the law: stated[5,33]	Same side. Defended its two red lines in court, with mixed results;[6,132] its products are being removed from defence systems,[80] while a civilian agency plans a bridge contract[146]	High	Defence revenue lost to the designation and the removal from defence systems; civilian agencies may continue under bridge contracts[3,80,146]	Safety first, Fairness first, Power spread. Would change: renegotiated defence terms
10. Should law require approval of agents' high-impact actions?	No stated position found	Leans to autonomy. Coding sessions are more automated than chat;[98] its framework asks approval before agents change systems. Users approve 93 per cent of permission requests; a separate test missed 17 per cent of 52 curated overeager actions. Neither statistic estimates unsafe actions in the population; high approval alone does not show failed oversight[12]	Low	A prior-approval rule would slow agentic use, much of which automates tasks; errors by agents carry liability[12,22]	None. Would change: a stated position on law

Axis and question	Anthropic's side and basis	Practice	Confidence	Financial route for Anthropic	Archetypes on the same side; what would change it
11. Should models obey operators, or may developers build in refusals?	Developer refusals: stated[60]	Same side. Constitution lets the model refuse, Anthropic included[60]	High	Refusals seen as overreach could cost customers, notably governments that want all lawful uses; no loss tied to model refusals is on the record	Safety first, Fairness first. Would change: a change to the constitution
12. Should marking of AI output be mandatory?	Mandatory: judgement, read from practice[14,75,76]	Same side. Watermarks text from newer models worldwide;[14] requires chatbots and external-facing agents to disclose that they are AI, a duty already in the September 2025 policy that the update of 12 November retains;[33,34] no statement on laws elsewhere found	Moderate	Compliance cost under the EU marking duty; the company says it watermarks newer models to comply[14,131]	Safety first, Fairness first, Self-reliance. Would change: a statement on marking laws
13. May developers train on creators' work without consent?	Use without consent: stated in litigation[2]	Same side. Music publishers, Canadian news publishers and Reddit suing;[19–21] no publisher licence reported[18]	High	Settlements and damages: a $1.5bn settlement of piracy claims, with suits by music and news publishers and Reddit pending[2,4,19–21]	Security first, Growth first, Self-reliance, Capital discipline. Would change: licensing or an opt-in
14. Who should bear the cost of the transition?	The firms: stated[23,25,49,96,147]	Same side, delivery incomplete. The Fund, a wholly owned subsidiary, has received 846,840,960 Class N shares, 5 per cent fully diluted assuming Amazon conversion, naming labour-market effects; the board controls it until its governance is approved; governance, liquidity and payouts remain separate;[26,52] proposals for worker support;[25] a $200m fund for outside research[24]	Moderate	The cost of a levy or worker support, and the contributed Fund equity; future additions of up to 2 per cent are discretionary at successive 50 per cent price milestones, requiring the threshold on 90 of 120 trading days and the preceding ten-day average. Future dilution is not automatic[25,52,147]	Safety first, Fairness first, Power spread. Would change: fund governance and payouts
15. Should AI build-out stay within a carbon budget?	Emissions may rise: inferred[37]	Same side. No company-wide emissions figure identified in the publications examined, including its first Public Benefit Report;[35,83] a gas-engine campus reported to be built for it, with potential emissions of about 11 million tons of CO2-equivalent a year[38,39]	Moderate	Permits, grid rules and local opposition can delay contracted capacity; carbon rules could raise its cost[39,82,133,148]	Security first, Growth first, Power spread, Self-reliance, Capital discipline. Would change: a carbon limit or disclosure

B4. How we read its policy positions against the scenarios

We ask two descriptive questions of a set of positions. First, does it favour restraint on the spread of AI? We count a closed frontier, a power to block dangerous deployments, restricted compute and slower progress (questions 1 to 4). Second, does it place duties on AI firms? We count mandatory and published disclosure, enforced competition, limits on defence use beyond the law, prior approval of agents' high-impact actions, developer refusals, mandatory marking of AI output, creators' consent to training, firms bearing the cost of displacement and a carbon budget for data centres (questions 5 and 8 to 15). Questions 6 and 7 count towards neither. Each placed answer counts equally, and unplaced answers are left out.

Applied to Anthropic, restraint is 4 of 4. Duties are 5 of 8, with one unplaced: disclosure, defence use, refusals, marking and the cost of displacement count towards duties (questions 5, 9, 11, 12 and 14); concentration, training without consent and emissions do not (questions 8, 13 and 15); agents are unplaced (question 10). The restraint count reads the September pacing proposal as qualified by the filing, which applies it chiefly to standards for the whole industry rather than to the company's own timetable.[49,55]

Together the two counts point to Luddite, as Exhibit 6 shows. The duties count is close: reading developer refusals as raising risk gives a tie at 4 of 8, and leaving marking unplaced gives 4 of 7, in which case the reading lies between Luddite and Icarus. Luddite as an outcome would also need rules that bind once commitments are hard to adjust, losses that adjustment, demand or a protected position do not offset, and controls that keep harm low. The tally describes positions. It does not forecast the scenario or measure the effects of the rules.

ANNEX C

The company's figures and contracts

The registration statement on Form S-1, filed on 26 October 2026, is the source for this annex.[1] Exhibit C1 gives the accounts and Exhibit C2 the named compute contracts. The filing gives minimum payments by year only for the commitments it held at 31 December 2025, which totalled $54.6bn. Its updated table of contractual obligations is blank.[1,27]

EXHIBIT C1

Selected figures from the registration statement ($m)

Audited for 2023 to 2025; unaudited for the six months to 30 June. Adjusted operating income is the company's own measure, not a figure under US generally accepted accounting principles. Costs in 2025 include a $1.5bn settlement in Bartz v. Anthropic. Most of the net loss in the first half of 2026 ($80.7bn) is the change in the fair value of convertible notes and similar instruments ($72.4bn); the operating loss was $0.7bn. Other expense of $8.8bn was mainly the initial recognition and later change in fair value of assets and liabilities under its convertible note purchase agreements, the filing says.

	2023	2024	2025	H1 2025	H1 2026
Revenue	41	386	4,585	1,163	16,286
Compute and infrastructure	433	2,522	7,326	2,334	8,945
Personnel	95	454	1,839	538	3,152
Sales, marketing and partnerships	2	43	568	120	3,168
Operating loss	(551)	(2,980)	(8,060)	(2,135)	(693)
Change in fair value of convertible notes and similar instruments	(1,045)	(5,434)	(34,308)	(10,942)	(72,388)
Net loss	(1,531)	(8,309)	(41,965)	(12,969)	(80,748)
Adjusted operating income (loss)	(522)	(2,788)	(4,854)	(1,849)	1,618

EXHIBIT C2

The compute book, from the registration statement

Amounts are as the filing states them: “at least” marks a minimum spend and “up to” a ceiling.

Counterparty	Terms	Term	Can it be reduced?	Notes
Google	At least about $114.5bn of spending, payable regardless of use; any shortfall paid to Google as a true-up	April 2026 to July 2033	No, except for Google’s uncured material breach	28
Google and other TPU providers	Further TPU capacity, estimated at $125bn to $150bn; terms not final; not in the table of commitments	2028 to 2031	Payments depend on capacity delivered and deployment conditions	28
Broadcom	TPU equipment leases, accounted for as finance leases, with rent of about $161.2bn excluding interest; up to $42.0bn of convertible notes may be issued to finance rent under the 2026 lease	Five-year master leases; rent normally starts when equipment is ready for service, but may start before facility readiness if construction is delayed	Non-cancellable except default; parent guarantees and cross-defaults may accelerate obligations	28
Amazon	At least $110.0bn of spending, payable regardless of use, with a true-up	May 2026 to April 2036	No, except for Amazon’s uncured material breach	28
Microsoft	At least $44.9bn of spending, payable regardless of use	November 2026 to May 2033	No, except for Microsoft’s uncured material breach or, tranche by tranche, for undelivered tranches	28
SpaceXAI	Up to $120.5bn, about $3.1bn a month once all capacity is delivered, which the company expects in December 2026	Into 2029	Either party may cancel on 90 days’ notice after non-cancellable periods ending in January and March 2027	28
AMD	No fixed dollar commitment; more than $20bn if the initial capacity is bought; not in the table of commitments	From 2027	No, except for AMD’s uncured material breach	28
Nscale	Up to about $44.6bn, with fees only as capacity is delivered and accepted; Nscale must use best efforts to obtain financing	Into 2033	Only in limited circumstances, including late delivery; step-in and replacement arrangements, as the parties agree, if Nscale defaults on its financing	28, 149
Lambda	About $41.3bn, with fees only as capacity is delivered; either party may end the August agreement, about $34.9bn, if Lambda does not secure the facility lease	Into 2034	Otherwise only for uncured breach or other limited circumstances	28
Fluidstack and others	**Signed:** a colocation and services agreement with Fluidstack; a $2.5bn escrow deposit for a compute lease; a guarantee of up to $500m of a provider’s power purchase obligations. **Not in the filing; from press reports and earlier statements:** arrangements with CoreWeave, Hut 8, TeraWulf, Volta, EverWind and NTT Data	Various	Not described	115, 123, 150–153

Notes

Dates are of publication, or of reading where a page carries no date. Press reports without a stable link are cited by outlet and date.

1 Anthropic, registration statement on Form S-1, filed with the US Securities and Exchange Commission on 26 October 2026: Cover; Prospectus summary; Use of proceeds; Principal and selling stockholders; Certain relationships and related person transactions. https://www.sec.gov/Archives/edgar/data/2130357/000162828026067168/anthropicpbc-sx1.htm. MD&A in later notes is the filing's management's discussion and analysis.

2 Order on Fair Use, Bartz v. Anthropic PBC, No. C 24-05417 WHA (N.D. Cal.), Dkt. 231, US District Court, Northern District of California (via CourtListener), 23 June 2025. https://storage.courtlistener.com/recap/gov.uscourts.cand.434709/gov.uscourts.cand.434709.231.0_2.pdf

3 Anthropic PBC v. United States Department of War, No. 26-1049 (D.C. Cir. Sept. 25, 2026), opinion, U.S. Court of Appeals for the D.C. Circuit, 25 September 2026. https://media.cadc.uscourts.gov/opinions/docs/2026/09/26-1049-2194984.pdf

4 Anthropic, registration statement (note 1): Risk factors; Notes, commitments and contingencies.

5 Statement from Dario Amodei on our discussions with the Department of War, Anthropic, 26 February 2026. https://www.anthropic.com/news/statement-department-of-war

6 Order on Cross Motions for Summary Judgment, Anthropic PBC v. U.S. Department of War, No. 26-cv-01996-RFL, Dkt. 250 (with Order of Final Relief, Dkt. 251), US District Court, Northern District of California (Judge Rita F. Lin), via CourtListener, 27 August 2026. https://storage.courtlistener.com/recap/gov.uscourts.cand.465515/gov.uscourts.cand.465515.250.0.pdf

7 Anthropic, "Activating AI Safety Level 3 protections", https://www.anthropic.com/news/activating-asl3-protections (22 May 2025)

8 Anthropic, Redacted Risk Report, August 2026, https://www.anthropic.com/aug-2026-risk-report (14 Aug 2026)

9 Anthropic, https://www.anthropic.com/glasswing; https://www.anthropic.com/news/cyber-verification-program (7 Apr and 6 Oct 2026)

10 Anthropic, "Disrupting the first reported AI-orchestrated cyber espionage campaign", https://www.anthropic.com/news/disrupting-AI-espionage (13 Nov 2025)

11 Anthropic, https://www.anthropic.com/research/alignment-assessment-ukaisi; AISI, https://www.aisi.gov.uk/blog/investigating-alignment-failures-in-aisis-incident (23 Oct 2026)

12 Anthropic, https://www.anthropic.com/news/our-framework-for-developing-safe-and-trustworthy-agents; https://www.anthropic.com/engineering/claude-code-auto-mode (4 Aug 2025; 25 Mar 2026)

13 Claude Help Center, "Can Claude produce images?" (16 Mar 2026), https://support.claude.com/en/articles/9002504-can-claude-produce-images (Read 24 Oct 2026)

14 How Claude's text watermark works, Anthropic, 14 August 2026. https://www.anthropic.com/news/claude-text-watermark

15 Anthropic, "Detecting and countering misuse of AI: September 2026", https://www.anthropic.com/threat-intelligence-report-september-2026 (10 Sep 2026)

16 Anthropic, "An update on our election safeguards", https://www.anthropic.com/news/election-safeguards-update; "Preparing for the 2026 US midterms", https://www.anthropic.com/news/2026-us-midterms (24 Apr and 14 Oct 2026)

17 Claude Help Center, crawler article (read 20 Oct 2026); Cloudflare, "The crawl-to-click gap", https://blog.cloudflare.com/crawlers-click-ai-bots-training/; "2025 Radar Year in Review", https://blog.cloudflare.com/radar-2025-year-in-review/ (29 Aug and 15 Dec 2025)

18 Adweek, https://www.adweek.com/media/anthropic-content-licensing-lawsuits-publishers/ (c. 27 Aug 2026)

19 N.D. Cal. docket 5:24-cv-03811; Music Business Worldwide, 29 Aug 2026 (28 Aug 2026)

20 CourtListener, https://www.courtlistener.com/docket/75002159/; CBC, https://www.cbc.ca/news/business/anthropic-postmedia-torstar-globe-canadian-press-copyright-lawsuit-9.7387205; Chat GPT Is Eating the World, 23 Oct 2026 (6 to 23 Oct 2026)

21 Crowell client alert; National Law Review, 24 Sep 2026 (4 Jun 2025 to Sep 2026)

Silicon Boundaries

22 Anthropic Economic Index: Tracking AI's role in the US and global economy (third Economic Index report), Anthropic, 15 September 2025. https://www.anthropic.com/research/economic-index-geography

23 Labor market impacts of AI: A new measure and early evidence, Anthropic, 5 March 2026. https://www.anthropic.com/research/labor-market-impacts

24 Anthropic, https://www.anthropic.com/news/introducing-the-anthropic-economic-futures-program; https://www.anthropic.com/news/economic-futures-research-fund-agenda (27 Jun 2025; 22 Jul 2026)

25 Anthropic, Economic Policy Framework (PDF), https://www-cdn.anthropic.com/files/4zrzovbb/website/9ea607a5dd67c168093829b701f3a0a6d21156d5.pdf (Jun 2026)

26 Anthropic, "Investing in the public benefit", 18 October 2026, https://www.anthropic.com/news/public-benefit-fund (checked 19 October 2026)

27 Anthropic, registration statement (note 1): Annual Note 16; interim Note 17; MD&A contractual obligations; Note 13 (purchase commitments at 31 December 2025).

28 Anthropic, registration statement (note 1): MD&A, infrastructure service arrangements and lease arrangements; Risk factors; pp171-173.

29 Amazon.com, Inc. Form 10-Q for the quarter ended June 30, 2026, Amazon.com, Inc. (SEC EDGAR), 31 July 2026. https://www.sec.gov/Archives/edgar/data/1018724/000101872426000026/amzn-20260630.htm

30 I. Progress and peril, BIS Annual Economic Report 2026, Bank for International Settlements, 28 June 2026. https://www.bis.org/publ/arpdf/ar2026e1.htm

31 Anthropic, registration statement (note 1): Summary consolidated financial data; MD&A, results of operations.

32 Anthropic, registration statement (note 1): Prospectus summary.

33 Anthropic, "2026 Usage Policy update", 8 October 2026, https://www.anthropic.com/news/2026-usage-policy-update; Usage Policy, https://www.anthropic.com/legal/aup

34 Anthropic, Usage Policy (archived version), https://anthropic.com/legal/archive/22742366-2ef0-4c7a-a833-6523f10d3944 (15 Sep 2025)

35 Stanford HAI, https://hai.stanford.edu/news/transparency-in-ai-is-on-the-decline; Heatmap News, 17 Jun 2026; Frontier Climate, https://frontierclimate.com/writing/growth-amc (9 Dec 2025; 17 Jun 2026)

36 Anthropic, registration statement (note 1): Risk factors p36.

37 Anthropic, "Build AI in America", 21 July 2025, https://www.anthropic.com/news/build-ai-in-america

38 Nscale and Microsoft announce collaboration with NVIDIA and Caterpillar to deliver 1.35GW of NVIDIA Vera Rubin NVL72 GPUs at flagship AI factory campus in West Virginia, Nscale (PR Newswire), 16 March 2026. https://www.prnewswire.com/news-releases/nscale-and-microsoft-announce-collaboration-with-nvidia-and-caterpillar-to-deliver-1-35gw-of-nvidia-vera-rubin-nvl72-gpus-at-flagship-ai-factory-campus-in-west-virginia-302715210.html

39 WVDEP, https://dep.wv.gov/news/Pages/Public-Comment-Period-and-Public-Participation-Opportunities-for-Monarch-Cloud-Campus.aspx (Notice of 29 Sep 2026)

40 Covering electricity price increases from our data centers, Anthropic, 11 February 2026. https://www.anthropic.com/news/covering-electricity-price-increases

41 Anthropic, registration statement (note 1): Prospectus summary; Business.

42 ChatGPT tops the list of chatbots Americans use, across age groups (from 'Americans and AI 2026: Chatbots, Smart Devices and Views on Impact'), Pew Research Center, 17 June 2026. https://www.pewresearch.org/chart/chatgpt-tops-the-list-of-chatbots-americans-use-across-age-groups/

43 Anthropic, Consumer Terms, https://www.anthropic.com/legal/consumer-terms; Claude Help Center, "Age assurance on Claude" (8 Oct 2025; 18 May 2026)

44 Anthropic, "Claude is a space to think", 4 February 2026, https://www.anthropic.com/news/claude-is-a-space-to-think; CNBC, 4 February 2026, on the ad-free pledge.

45 Anthropic, "Protecting the wellbeing of our users", https://www.anthropic.com/news/protecting-well-being-of-users (18 Dec 2025)

46 Anthropic, https://claude.com/resources/articles/claude-duet; https://www.anthropic.com/news/donating-agent-skills-to-the-agentic-ai-foundation; https://www.anthropic.com/research/formal-verification (20 to 23 Oct 2026)

47 Menlo Ventures (an Anthropic investor), https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/ (31 Jul and 9 Dec 2025)

48 UK Competition and Markets Authority, https://www.gov.uk/cma-cases/amazon-slash-anthropic-partnership-merger-inquiry and https://www.gov.uk/cma-cases/alphabet-inc-google-llc-slash-anthropic-merger-inquiry; Federal Trade Commission, "FTC Issues Staff Report on AI Partnerships & Investments Study", 17 January 2025, https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-issues-staff-report-ai-partnerships-investments-study (Sep 2024 to Jan 2025)

49 Dario Amodei, "We Must Pace the Frontier", 12 September 2026, https://darioamodei.com/post/we-must-pace-the-frontier

50 Anthropic, “Our position on open-weights models”, 27 July 2026, https://www.anthropic.com/news/position-open-weights-models

51 The Token State of Play (Inferred, issue 1), OpenRouter, 22 September 2026. https://openrouter.ai/data/newsletter/the-token-state-of-play

52 Anthropic, registration statement (note 1): MD&A, Our Commitment to Public Benefit, pp144-145; governance section.

53 Anthropic, “An update on independent evaluators”, 15 October 2026, updated 16 October, https://www.anthropic.com/news/an-update-on-independent-evaluators (checked 19 October 2026)

54 METR, https://metr.org/blog/2026-10-20-misalignment-risk-report-august-2026-review/; SecureBio, https://securebio.org/blog/response-to-anthropics-update-on-independent-evaluators/ (16 and 20 Oct 2026)

55 Anthropic, registration statement (note 1): Business, safety, p234; risk factors p52.

56 Office of the New York City Comptroller, “Letter to Anthropic Re: Planned Initial Public Offering”, 14 October 2026, https://comptroller.nyc.gov/reports/letter-to-anthropic-re-planned-initial-public-offering/ (read 19 October 2026); talks reported by Crain’s New York Business, 16 October 2026.

57 Anthropic, registration statement (note 1): Risk factors.

58 Anthropic, registration statement (note 1): Risk factors; Business, legal proceedings; Notes.

59 D.C. Cir. order, https://storage.courtlistener.com/recap/gov.uscourts.cadc.42923/gov.uscourts.cadc.42923.1208901294.0.pdf; docket, https://www.courtlistener.com/docket/72380208/anthropic-pbc-v-united-states-department-of-war/; DefenseScoop, https://defensescoop.com/2026/10/21/dc-circuit-decision-anthropic-pentagon-fight/ (20 to 23 Oct 2026)

60 Anthropic, Claude’s Constitution, https://www.anthropic.com/constitution; OpenAI, Model Spec, 18 August 2026, https://model-spec.openai.com/2026-08-18.html

61 Silicon Boundaries, “What do you believe about AI?”, research note, 29 October 2026, https://sustainablefinance.ox.ac.uk/wp-content/uploads/2026/10/2026.10.29_AI-Beliefs-Research-Note_Designed_final.pdf

62 Anthropic’s Responsible Scaling Policy: Version 3.0, Anthropic, 24 February 2026. https://www.anthropic.com/news/responsible-scaling-policy-v3

63 Anthropic, registration statement (note 1): MD&A, key factors.

64 Anthropic, registration statement (note 1): Our governance structure, pp245-248; Management, pp256-257; description of capital stock.

65 Anthropic, registration statement (note 1): Prospectus summary; Risk factors.

66 Introducing Claude Fable 5.1 and Claude Mythos 5.1 (Cost and availability), Anthropic, 1 September 2026. https://www.anthropic.com/claude-fable-and-mythos-5-1

67 Scoop: Anthropic breaks from major tech group over chips, Axios (Maria Curi), 8 September 2026. https://www.axios.com/2026/09/08/anthropic-breaks-tech-group-chips

68 Anthropic PBC v. U.S. Department of War, 3:26-cv-01996 (N.D. Cal.) docket, CourtListener (US District Court, Northern District of California), 9 March 2026. https://www.courtlistener.com/docket/72379655/anthropic-pbc-v-us-department-of-war/

69 Investigating three real-world incidents in our cybersecurity evaluations, Anthropic, 30 July 2026. https://www.anthropic.com/news/investigating-incidents-cybersecurity-evals

70 Anthropic expands partnership with Google and Broadcom for multiple gigawatts of next-generation compute, Anthropic, 6 April 2026. https://www.anthropic.com/news/google-broadcom-partnership-compute

71 Anthropic’s Annualized Revenue Tops $65 Billion Before IPO, Bloomberg News (via Bloomberg Industry Group), 18 August 2026. https://news.bgov.com/legal-ops-and-tech/anthropic-revenue-run-rate-surpasses-65-billion-ahead-of-ipo

72 September 2026 Ramp AI Index: Cracks in the AI thesis, part 2, Ramp Economics Lab (Ara Kharazian), 9 September 2026. https://ramp.com/data/ai-index-sept-2026

73 Policy on the AI Exponential (Advanced AI Framework), Anthropic, 10 June 2026. https://www.anthropic.com/policy-on-the-ai-exponential

74 Statement on the US government directive to suspend access to Fable 5 and Mythos 5, Anthropic, 12 June 2026. https://www.anthropic.com/news/fable-mythos-access

75 The General-Purpose AI Code of Practice (signatories list), European Commission, AI Office (Shaping Europe’s digital future), 10 July 2025 (code published; signatory list live and checked 19 October 2026). https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

76 Strong backing for the Code of Practice on Transparency of AI-generated Content, European Commission, AI Office (Shaping Europe’s digital future), 31 July 2026. https://digital-strategy.ec.europa.eu/en/news/strong-backing-code-practice-transparency-AI-generated-content

77 Anthropic, “Anthropic opens Singapore office and signs MOU with the Government of Singapore”, 14 October 2026, https://www.anthropic.com/news/singapore-office-opening

78 US lifts restrictions on Anthropic's powerful AI models Fable and Mythos, Al Jazeera (John Power), 1 July 2026. https://www.aljazeera.com/amp/economy/2026/7/1/us-lifts-restrictions-on-powerful-ai-models-fable-mythos-anthropic-says

79 Anthropic, registration statement (note 1): Risk factors; MD&A, liquidity and capital resources, pp162-163.

80 Congressional Research Service, "Federal Government and Anthropic", IF13217, https://www.congress.gov/crs_external_products/IF/PDF/IF13217/IF13217.4.pdf (14 Oct 2026)

81 Anthropic, "Detecting and preventing distillation attacks", https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks (23 Feb 2026)

82 FERC Launches Aggressive Targeted Action to Speed Large Load Integration, Federal Energy Regulatory Commission, 18 June 2026. https://www.ferc.gov/news-events/news/ferc-launches-aggressive-targeted-action-speed-large-load-integration

83 Anthropic, 2026 Public Benefit Report, https://www-cdn.anthropic.com/a9ceec76a54d28d27affffcdab6faf7c142cd587.pdf (21 Oct 2026)

84 CNBC, https://www.cnbc.com/2026/10/20/anthropic-nscale-data-center-deal.html; Nscale, Form S-1, https://www.sec.gov/Archives/edgar/data/0002110365/000119312526395475/ck0002110365-20260918.htm (19 to 22 Oct 2026; 18 Sep 2026)

85 Anthropic, https://www.anthropic.com/news/claude-opus-4-5; https://www.anthropic.com/news/claude-haiku-5-5; pricing page, read 20 Oct 2026 (24 Nov 2025; 7 Oct 2026)

86 Anthropic, https://www.anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation (9 Dec 2025)

87 CNBC, https://www.cnbc.com/2026/10/19/coreweave-oracle-ai-credit-stress.html and https://www.cnbc.com/2026/10/20/coreweave-debt-janus-henderson-lloyd.html; FT via Hedgeweek, https://hedgeweek.com/news/ai-debt-concerns-fuel-surge-in-credit-default-swap-trading (15 to 22 Oct 2026)

88 Anthropic, "Detecting and countering malicious uses of Claude: March 2025", https://www.anthropic.com/news/detecting-and-countering-malicious-uses-of-claude-march-2025 (23 Apr 2025)

89 Anthropic, "Measuring political bias in Claude", https://www.anthropic.com/news/political-even-handedness (13 Nov 2025)

90 Anthropic, "How people ask Claude for personal guidance", https://www.anthropic.com/research/claude-personal-guidance (30 Apr 2026)

91 Anthropic, "Detecting and countering misuse of AI: August 2025", https://www.anthropic.com/news/detecting-countering-misuse-aug-2025 (27 Aug 2025)

92 Anthropic, "How people use Claude for support, advice, and companionship", https://www.anthropic.com/news/how-people-use-claude-for-support-advice-and-companionship (27 Jun 2025)

93 AI Lawsuit Tracker, https://ailawsuittracker.com/companies/anthropic/ (Reviewed 19 Oct 2026)

94 AP via NBC Bay Area, 30 Sep 2026; Axios, 15 Oct 2026 (30 Sep and 15 Oct 2026)

95 Anthropic, "Disempowerment patterns in real-world AI usage", https://www.anthropic.com/news/disempowerment-patterns (28 Jan 2026)

96 Economic Scenarios for Transformative AI (Anthropic Institute Working Paper No. 2026-02), Anthropic Institute (Korinek, Jones, Sacher, Cotter, McCrory), September 2026. https://www-cdn.anthropic.com/files/4zrzovbb/website/cf58f84d46a4a76bf5a5b039ac695fba6b80041c.pdf

97 Anthropic, https://www.anthropic.com/research/anthropic-economic-index-september-2025-report; https://www.anthropic.com/research/anthropic-economic-index-january-2026-report (15 Sep 2025; 15 Jan 2026)

98 Anthropic Economic Index report: Cadences, Anthropic, 26 June 2026. https://www.anthropic.com/research/economic-index-june-2026-report

99 Tracking the Impact of AI on the Labor Market, The Budget Lab at Yale, 15 September 2026. https://budgetlab.yale.edu/research/tracking-impact-ai-labor-market

100 Yale Budget Lab, https://budgetlab.yale.edu/research/tracking-impact-ai-labor-market; Pew, https://www.pewresearch.org/short-reads/2026/10/22/men-are-experiencing-more-job-growth-than-women-in-ai-exposed-occupations/ (14 and 22 Oct 2026)

101 Ramp AI Index, https://ramp.com/data/ai-index; OpenRouter, https://openrouter.ai/rankings (Viewed 24 Oct 2026)

102 Anthropic, https://www.anthropic.com/news/introducing-claude-for-education; AFT, https://www.aft.org/node/101397; Anthropic, https://www.anthropic.com/news/claude-for-teachers; Claude Help Center, "Set up Claude for Teachers for your school or district", https://support.claude.com/en/articles/16559896-set-up-claude-for-teachers-for-your-school-or-district (Apr 2025 to Jul 2026)

103 Anthropic, https://www.anthropic.com/news/gates-foundation-partnership (14 May 2026)

104 Anthropic, https://www.anthropic.com/news/claude-for-life-sciences; https://www.anthropic.com/news/healthcare-life-sciences (20 Oct 2025; 11 Jan 2026)

105 Business Wire, 7 Nov 2024; Anthropic, https://www.anthropic.com/news/claude-gov-models-for-u-s-national-security-customers; https://www.anthropic.com/news/anthropic-and-the-department-of-defense-to-advance-responsible-ai-in-defense-operations (Nov 2024 to Aug 2025)

106 Anthropic, "What we learned mapping a year's worth of AI-enabled cyber threats", https://www.anthropic.com/news/AI-enabled-cyber-threats-mitre-attack (3 Jun 2026)

107 Anthropic, Responsible Scaling Policy 3.2 and 3.3, https://www.anthropic.com/responsible-scaling-policy (26 May 2026)

108 Anthropic Responsible Scaling Policy, Version 3.3 (changelog entry May 26, 2026), Anthropic, 26 May 2026. https://www-cdn.anthropic.com/files/4zrzovbb/website/c11e84981d0a7281a1b229f3fa6af0da66eaf43f.pdf

109 Anthropic, https://www.anthropic.com/news/updating-restrictions-of-sales-to-unsupported-regions (4 Sep 2025)

110 Anthropic, https://www.anthropic.com/research/prompt-injection-defenses; Claude Sonnet 5.5 system card (24 Nov 2025; 28 Sep 2026)

111 Bloomberg, "Anthropic's Q3 revenue jumps to $16.8 billion ahead of planned IPO", 15 October 2026, via Business Standard, https://www.business-standard.com/companies/quarterly-results/anthropic-s-q3-revenue-jumps-to-16-8-billion-ahead-of-planned-ipo-126101501725_1.html (people familiar with the matter; the company declined to comment)

112 Anthropic, https://www.anthropic.com/news/claude-for-financial-services; https://www.anthropic.com/news/barclays-scales-claude (15 Jul 2025; 1 Oct 2026)

113 FSB, https://www.fsb.org/uploads/P101025.pdf; Bank of England, Financial Stability Report, July 2026 (Apr 2025 to Jul 2026)

114 Anthropic raises $65B in Series H funding at $965B post-money valuation, Anthropic, 28 May 2026. https://www.anthropic.com/news/series-h

115 Hut 8 Signs 15-Year, 245 MW AI Data Center Lease at River Bend Campus with Total Contract Value of $7.0 Billion (Form 8-K, Exhibit 99.1), Hut 8 Corp. (SEC EDGAR), 17 December 2025. https://www.sec.gov/Archives/edgar/data/1964789/000110465925122052/hut-20251217xex99d1.htm

116 IMF Blog, https://www.imf.org/en/blogs/articles/2026/10/12/global-financial-markets-face-rising-risks-from-ai-bond-yields-and-cyberattacks (13 Oct 2026)

117 FSB, https://www.fsb.org/2026/10/sound-practices-for-responsible-adoption-of-artificial-intelligence-ai-final-report/; https://www.fsb.org/2026/10/innovation-and-resilience-managing-the-risks-of-a-changing-financial-system/ (15 and 21 Oct 2026)

118 Anthropic, "Updates to Consumer Terms and Privacy Policy", https://www.anthropic.com/news/updates-to-our-consumer-terms; Privacy Center (28 Aug 2025; 10 Sep and 14 Oct 2026)

119 Courthouse News, 30 Sep 2026; Claude Help Center, "Identity verification on Claude" (30 Sep 2026)

120 Anthropic, "Evaluating and Mitigating Discrimination in Language Model Decisions", https://www.anthropic.com/research/evaluating-and-mitigating-discrimination-in-language-model-decisions (7 Dec 2023)

121 Anthropic and Amazon expand collaboration for up to 5 gigawatts of new compute, Anthropic, 20 April 2026. https://www.anthropic.com/news/anthropic-amazon-compute

122 Higher usage limits for Claude and a compute deal with SpaceX, Anthropic, 6 May 2026. https://www.anthropic.com/news/higher-limits-spacex

123 AI cloud startup Volta valued at $2.4 billion, announces $10 billion AI partnership (Reuters), Reuters, via The Star (Malaysia), 4 August 2026. https://www.thestar.com.my/tech/tech-news/2026/08/04/ai-cloud-startup-volta-valued-at-24-billion-announces-10-billion-ai-partnership

124 Form S-1, Space Exploration Technologies Corp. (Prospectus Summary: compute services agreements), SpaceX (SEC EDGAR), 20 May 2026. https://www.sec.gov/Archives/edgar/data/1181412/000162828026036936/spaceexplorationtechnologi.htm

125 N.D. Miss. No. 3:26-cv-74, order of 19 Oct 2026; MDEQ Agreed Order No. 7739 26 (30 Jul and 19 Oct 2026)

126 AMD and Anthropic Announce Strategic Partnership to Deploy Up to 2 Gigawatts of AMD Instinct MI450 Series GPUs, AMD, 22 July 2026. https://ir.amd.com/news-events/press-releases/detail/1292/amd-and-anthropic-announce-strategic-partnership-to-deploy-up-to-2-gigawatts-of-amd-instinct-mi450-series-gpus

127 Form 8-K, Broadcom Inc. (Item 8.01, Other Events, event date 6 April 2026), Broadcom Inc., via SEC EDGAR, 6 April 2026. https://www.sec.gov/Archives/edgar/data/1730168/000119312526144028/d87999d8k.htm

128 Dario Amodei, "On DeepSeek and Export Controls", https://www.darioamodei.com/post/on-deepseek-and-export-controls (Jan 2025)

129 Anthropic, registration statement (note 1): MD&A, liquidity.

130 California Business and Professions Code section 22757.13 (as added by SB 53, Transparency in Frontier Artificial Intelligence Act), California Legislature (statute text as mirrored by public.law, updated 1 Jan 2026), 29 September 2025. https://california.public.law/codes/business_and_professions_code_section_22757.13

131 Regulation (EU) 2024/1689 (the AI Act), Article 50, https://eur-lex.europa.eu/eli/reg/2024/1689/oj; on the AI Omnibus and the deadline of 2 December 2026 for systems already on the market, Kingsley Napley, 22 July 2026.

132 U.S. appeals court upholds Pentagon designation of Anthropic as supply chain risk, CNBC, 25 September 2026. https://www.cnbc.com/2026/09/25/pentagon-anthropic-ai-risk-appeals-court.html

133 Spectrum News, 24 Aug 2026 (24 Aug 2026)

134 Anthropic Responsible Scaling Policy, Version 3.4 (changelog entry July 8, 2026), Anthropic, 8 July 2026. https://www-cdn.anthropic.com/files/4zrzovbb/website/0bacdc8440ea96e62a8766d99ebe1d4eea6d5f3a.pdf

135 Form S-1 registration statement, Nscale Ltd (filed 18 September 2026), Nscale Ltd, via SEC EDGAR, 18 September 2026. https://www.sec.gov/Archives/edgar/data/2110365/000119312526395475/ck0002110365-20260918.htm

136 TechCrunch, 2 September 2026, on OpenAI's recurrent depth; OpenAI, "GPT-6 Astra System Card", section 9, "Monitorability", 3 September 2026, https://deploymentsafety.openai.com/gpt-6-astra; Shah and Dragan, "The case for reasoning transparency", DeepMind Institute, 16 September 2026.

137 Hyperscaler debt binge pushes yields up as investor demand cools, Reuters (syndicated by WDEZ), 29 July 2026. https://wdez.com/2026/07/29/hyperscaler-debt-binge-pushes-yields-up-as-investor-demand-cools/

138 Reuters via Al Jazeera (21 Oct) and via SRN News, https://srnnews.com/analysis-anthropic-ipo-stands-alone-as-ai-market-jitters-sideline-others/ (15 and 22 Oct 2026)

139 Anthropic Expands Credit Facility to $15 Billion Ahead of Mega IPO (PYMNTS, relaying Bloomberg), PYMNTS, 4 September 2026. https://www.pymnts.com/news/artificial-intelligence/2026/anthropic-expands-credit-facility-to-15-billion-ahead-of-mega-ipo/

140 "Open Weights and American AI Leadership", letter hosted by Nvidia, 24 July 2026: 25 signatories at launch, with OpenAI, Anthropic and Google absent (Tom's Hardware, 24 July 2026); more than 270 by 3 August 2026, among them OpenAI, Google and Amazon, as Microsoft's page records; Anthropic absent from that page and from the 235 names in the Nvidia-hosted PDF as read on 17 October 2026, which include SpaceX.

141 NIST, US AI Safety Institute agreements with Anthropic and OpenAI, 29 August 2024; NIST, update on the Center for AI Standards and Innovation's work with OpenAI and Anthropic, 25 September 2025, https://www.nist.gov/news-events/news/2025/09/caisi-works-openai-and-anthropic-promote-secure-ai-innovation; CIO Dive, 5 May 2026, on the Google DeepMind, Microsoft and xAI agreements.

142 Anthropic is endorsing SB 53, Anthropic, 8 September 2025. https://www.anthropic.com/news/anthropic-is-endorsing-sb-53

143 Sharing our compliance framework for California's Transparency in Frontier AI Act, Anthropic, 19 December 2025. https://www.anthropic.com/news/compliance-framework-SB53

144 Anthropic's Long-Term Benefit Trust appoints Vas Narasimhan to Board of Directors, Anthropic, 14 April 2026. https://www.anthropic.com/news/narasimhan-board

145 The Long-Term Benefit Trust, Anthropic, 19 September 2023. https://www.anthropic.com/news/the-long-term-benefit-trust

146 FedScoop, https://fedscoop.com/hhs-sole-source-award-anthropic-ai-claude/ and https://fedscoop.com/hhs-sole-source-award-to-anthropic-continues-to-raise-questions/ (19 to 23 Oct 2026)

147 Axios, 28 May 2025, on Amodei's proposed "token tax".

148 WSAZ, 14 Jul 2026; Charleston Gazette-Mail, c. 17 Oct 2026 (14 Jul and c. 17 Oct 2026)

149 Nscale, Form S-1, Exhibit 10.25 (GPU services agreement with Anthropic), filed 18 September 2026, https://www.sec.gov/Archives/edgar/data/2110365/000119312526395475/ck0002110365-ex10_25.htm

150 Anthropic, registration statement (note 1): Business, compute; Notes, subsequent events; MD&A.

151 CoreWeave Announces Multi-Year Agreement With Anthropic, CoreWeave, 10 April 2026. https://coreweave.com/news/coreweave-announces-multi-year-agreement-with-anthropic

152 Hut 8 Corp. Form 10-K for the fiscal year ended 31 December 2025, Hut 8 Corp. (SEC EDGAR), 25 February 2026. https://www.sec.gov/Archives/edgar/data/1964789/000110465926019392/hut-20251231x10k.htm

153 Reuters via SRN News, https://srnnews.com/anthropic-signs-lease-for-first-canadian-data-center/; Nikkei via The Star, https://www.thestar.com.my/tech/tech-news/2026/10/13/ntt-data-in-final-talks-to-build-anthropic039s-first-data-centre-in-japan-nikkei-reports (13 and 16 Oct 2026)

Silicon Boundaries

About Silicon Boundaries

Silicon Boundaries is a joint initiative of the Oxford Sustainable Finance Group at the University of Oxford and Theia Finance Labs. It asks how much AI society can productively absorb, and how to raise that amount over time, while staying within a safe operating space. Silicon Boundaries define that space. They are the thresholds beyond which the benefits of AI and related technologies are outweighed by their contribution to societal instability.

sustainablefinance.ox.ac.uk/research/silicon-boundaries/

Contact

Ben Caldecott
ben.caldecott@smithschool.ox.ac.uk

Jakob Thomä
jakob@theiafinance.org

THEIA FINANCE LABS

OXFORD SUSTAINABLE FINANCE GROUP