

Scoring Protocol

Pre-registered comparison of the Anthropic analogical risk inventory with the forthcoming Form S-1

Robert G. Eccles and Shivaram Rajgopal

Working Paper 3. Fixed before filing. Companion to Working Paper 1, "Anthropic Before the Prospectus" (information cutoff September 26, 2026). The predictions are fixed as of publication; any corrections appear only as dated errata. Items marked **[prediction]** are pre-registered forecasts.

1. What Is Scored, and What Is Not

This protocol fixes, before the prospectus exists, how our analogical risk inventory will be compared with Anthropic's Form S-1 once it is publicly filed. It is written against one governing premise. A risk-factor section is a legal artifact, not a materiality assessment. Counsel drafts it to be over-inclusive, orders it by convention, and leaves it silent where silence is preferred. Coverage in the prospectus is therefore never treated here as the ground truth of materiality. A risk we identified that the prospectus omits is not, by that fact alone, a risk we got wrong. A risk the prospectus lists that we did not identify is not, by that fact alone, a risk we missed through analytical failure.

The protocol therefore scores four objects, and each is protected against the premise in a different way. The first is coverage, running from the inventory to the prospectus (Section 3). It records how the prospectus treats each of the fifteen risks in the inventory. Its outcomes describe the prospectus, and they are scored against our pre-registered expectations, not against materiality. The second is capture, running from the prospectus to the inventory (Section 4). It records which of the prospectus's company-specific risk factors the inventory anticipated. Each miss is classified by a fixed rule into information asymmetry, design-induced omission or analytical blind spot, and only the last counts against the inventory. The third is a set of magnitude checks (Section 5), applied wherever the prospectus discloses a number that can be set against a pre-registered figure. The fourth is a set of pre-registered predictions (Section 6). These cover the answer to each of the twenty-six checkable questions and the coverage outcome for each inventory item, each with a stated probability, and they are scored by accuracy and Brier score. The predictions are the component on which we can be unambiguously wrong, and we regard them as the study's most demanding test.

2. Document, Version and Units

The scored document is the first registration statement on Form S-1 that Anthropic, PBC, or any successor issuer of its shares, files publicly on EDGAR. The confidential draft is not scored. All parts of the filed document are in scope: the prospectus summary, risk factors, management's discussion and analysis, business, legal proceedings, management and corporate governance, executive compensation, certain relationships and related-party transactions, description of capital stock, and the financial statements and notes. Amendments on Form S-1/A are scored as a second layer and reported only as changes relative to the first public filing, so that the result reflects what the issuer chose to say first. If no public registration statement has been filed by June 30, 2027, we will publish a report of non-occurrence, and the predictions will be scored as unresolved rather than as misses.

A risk factor is each separately captioned entry in the section headed Risk Factors. The caption and the text beneath it form one unit, even when the text discusses several mechanisms, and the unit takes the mechanism named in the caption as its principal risk. Before scoring capture, each coder removes

boilerplate risk factors. A risk factor is boilerplate if it falls within one of the following ten fixed categories and contains nothing specific to Anthropic's business, technology, counterparties or governance. The categories are: (i) volatility of the share price or absence of a prior market; (ii) dilution from future issuance; (iii) costs of being a public company, including internal-control certification; (iv) dividend policy; (v) generic anti-takeover and forum-selection provisions; (vi) analyst coverage; (vii) generic changes in tax law; (viii) macroeconomic conditions not tied to demand for AI; (ix) management's discretion over proceeds; and (x) generic exemptions available to controlled or emerging companies. Founder voting control, the Long-Term Benefit Trust and public-benefit-corporation duties are company-specific and are never boilerplate.

3. Coverage: From the Inventory to the Prospectus

Each inventory item receives one of three codes, assigned from anywhere in the scored document. A hit means the prospectus addresses the risk and identifies the mechanism the inventory identified; the minimum content for each item is fixed in the table below. A partial means the prospectus addresses the risk only generically, bundles it without its mechanism, locates it through a materially different mechanism, or discloses the underlying facts only in the financial-statement notes without discussing the risk. A miss means the risk is not addressed. Page references for every code are recorded. Coverage outcomes are reported item by item and compared with the predictions in Table 3. We do not compute a coverage "score" that would treat prospectus silence as our error.

#	Inventory item	Minimum content for a hit	Illustrative partial
1	Compute commitment mismatch	Identifies the risk that long-term compute commitments exceed realized demand, or that capacity falls short of demand, with reference to fixed or non-cancellable obligations	States that substantial compute is required without discussing commitment risk
2	Energy and host community	Identifies power availability, electricity-price or community opposition as a constraint on capacity, or discloses the electricity-price pledge	Generic data-center availability risk
3	Supplier–investor concentration	Identifies dependence on a small number of compute suppliers who are also shareholders or financing counterparties	Identifies supplier dependence without the investor relationship, or vice versa
4	Customer and channel concentration	Identifies concentration of revenue in a limited number of customers or distribution channels	Mentions reliance on partners for distribution only
5	Revenue quality and presentation	Reconciles run-rate or annualized metrics to GAAP revenue and states principal-versus-agent treatment for sales through cloud partners	Does one of the two
6	Sovereign access discretion	Identifies government restriction of access to models or chips, with reference to the June 2026 suspension or equivalent	Generic export-control or trade-policy risk
7	Procurement designation	Discloses the Department of War supply-chain-risk designation and its possible effect on commercial relationships	Generic risk of losing government business
8	Training-data acquisition liability	Identifies claims arising from how training data was acquired, as distinct from whether training is fair use, and discloses the music litigation	Generic statement that copyright claims may be brought
9	Output IP and customer indemnity	Discloses the obligation to indemnify customers against IP claims and that it sits outside ordinary liability limits	Mentions indemnification obligations without scope or cap
10	Harm from use	Identifies liability for harm to users from model outputs, including vulnerable users or minors	Generic reputational risk from harmful or inaccurate outputs
11	Dual-use misuse and uninsurability	Identifies misuse of model capabilities by third parties for cyber or biological harm and the exposure this creates	Generic misuse or abuse of services

#	Inventory item	Minimum content for a hit	Illustrative partial
12	Agentic deployment and third-party reliance	Identifies losses from autonomous actions taken by agentic products in customer environments, or from reliance by parties downstream of customers	Generic product defects or service errors
13	Regulatory proliferation	Identifies frontier-AI-specific regulation (state safety laws, EU AI Act or equivalents), including incident reporting obligations	Generic statement that AI regulation is evolving
14	Deployment restraint	Identifies that the company's own safety commitments may require it to delay, restrict or withhold models, at a commercial cost	States a commitment to safety without identifying its commercial cost
15	Risk-oversight architecture and control structure	Discusses founder voting control or Trust board-selection rights and their consequences for public holders, and describes board-level oversight of model risk	Discusses control structure without risk oversight, or the reverse

Table 1. Coverage anchors (fixed before filing)

4. Capture: From the Prospectus to the Inventory

For each non-boilerplate risk factor, the coders ask whether its principal risk is anticipated by the inventory. A hit means it maps to an inventory item with the same mechanism. A partial means it maps to an item through a different mechanism, or to a mechanism discussed in the inventory's cluster analysis but not carried into the consolidated table. A miss means neither. The raw capture rate is hits plus half of partials, divided by the number of non-boilerplate risk factors.

Every miss is then assigned to exactly one of three classes, tested in order. Class A is information asymmetry: the fact on which the risk factor rests was not public on or before September 26, 2026. The test is fixed. Each coder searches the source register, then runs three web searches restricted to material dated on or before the cutoff, using the risk factor's caption and its two most specific nouns. If neither coder finds a public source, the miss is Class A. Events occurring after the cutoff, such as a suit filed in October, form subclass A2 and are reported separately. Class B is design-induced omission: the risk falls within the five categories that the inventory declared, before filing, that its analogy set does not generate. These are price competition among frontier laboratories, commoditization by open-weight models, obsolescence of model generations, dependence on key personnel, and the security of model weights and source code. The list is closed and cannot be extended after publication. Class C is analytical blind spot: every miss that is neither A nor B. Each Class C miss is listed individually, with the public source dated on or before the cutoff that would have revealed it.

We report three figures: the raw capture rate; the adjusted capture rate, which removes Class A and Class B misses from the denominator; and the Class C count with its list. Only Class C counts against the inventory as analysis. Class B counts against the choice of analogies, which the study's authors made and own. Class A counts against nobody. It is the measure of what the prospectus added to public knowledge, and it is itself a finding about the informational value of an IPO filing for a company this heavily reported.

5. Magnitude Checks

Where the prospectus discloses a figure comparable to one of the figures pre-registered in Table 2, the figure is scored against our stated range. A hit means the figure falls within the range. A partial means it falls outside the range but within a factor of ten. A miss means it is further away. The comparables we expect to be available are fixed in Table 2. A figure not disclosed is recorded as unscorable, not as a miss.

Comparable	Our pre-registered figure or range	Basis
Revenue, nine months to September 30, 2026	\$29–38 billion <i>[prediction: first half \$15.7B reported or projected, plus a third quarter of about \$14–22B implied by run-rates between \$65B and roughly \$80B]</i>	Derived here from reported and projected quarterly revenue [S14, S15] and run-rates [S11]
Revenue, full year 2026 (if disclosed or guided)	\$50–60 billion <i>[prediction]</i>	Derived here from the first-half figures and the year-end run-rate projection [S06]
Bartz settlement: remaining obligation at the balance-sheet date	\$450 million plus interest	Inventory Section 3, Asbestos [S87]
Music litigation: accrual, or range of reasonably possible loss	No accrual; if a range is given, within \$0.06–3 billion	Inventory Section 3, Asbestos: bounded by the Bartz per-work rate and the statutory maximum
Total purchase, lease and compute commitments	\$250–550 billion <i>[prediction]</i>	Inventory Section 4, Telecommunications [S31]
Largest single customer or channel share of revenue	10–30 percent	Inventory Section 5, Semiconductors [S13]
Revenue effect of the June 2026 suspension	\$0.3–1.2 billion <i>[prediction]</i>	Derived here: both flagship models disabled for about three weeks; if they carried a third of usage at a \$60 billion run-rate, affected revenue was about \$1.2 billion, much of it migrating to older models
AI-specific liability insurance limits	Below \$1 billion, or self-insured	Inventory Section 3, Asbestos [S101]
Pre-tax income or loss, nine months to September 30, 2026	Between a loss of \$5 billion and a profit of \$3 billion <i>[prediction]</i>	Derived here from the reported first operating profit and the company's caution that it might not last [S14]

Table 2. Magnitude comparables (fixed before filing)

6. Pre-Registered Predictions

For each of the twenty-six checkable questions, we predict the category of answer the prospectus will give, using the answer forms fixed in the inventory's Table 2, and attach a probability that the predicted category occurs. Each prediction is scored as a binary event. The Brier component is $(p - o)^2$, where p is our stated probability and o is 1 if the predicted category occurs and 0 otherwise. We report the mean Brier score against the uninformative benchmark of 0.25 and the count of categorical hits. The predictions for coverage outcomes are scored identically. A caution applies to the model that drafted these predictions: it was trained heavily on Anthropic's self-description. Predictions about how Anthropic will describe its safety commitments (questions 6, 7, 13, 15 and 16, and coverage items 14 and 15) are the ones most exposed to recall, so we flag them and report their accuracy separately.

Q	Question (abbreviated)	Predicted answer	p	Basis
1	Music litigation: accrue or disclose	Disclose, no accrual	0.70	Fifty-fold spread between defensible bounds; ASC 450 "not estimable"
2	Offering as qualifying funding event for Bartz	No explicit statement	0.60	Payment terms likely disclosed without characterizing the trigger
3	Indemnification of officers named personally	No specific disclosure	0.60	Standard director and officer indemnity language expected, not case-specific
4	AI-specific insurance, captive or self-insurance	Described, not quantified	0.60	Insurance limits rarely quantified in risk factors

Q	Question (abbreviated)	Predicted answer	p	Basis
5	Incident system and SB 53 / RAISE report counts	Described, no counts	0.55	Laws named; counts would invite scrutiny
6	Revenue forgone from withheld or gated models *	Silent	0.60	Quantification would be forward-looking and competitively sensitive
7	Gating as a control whose failure is a risk *	Partial	0.50	Weight-security risk likely; gating failure framed as security, not liability
8	Intermediary immunity for output	Disclaim (says immunity may not apply)	0.45	Counsel likely to warn Section 230 may be unavailable
9	Age assurance incl. minors via API applications	Partial	0.50	Consumer age policy described; downstream API minors less likely
10	Own published research as litigation risk	No	0.70	Unusual for issuer to flag its own research
11	Liability caps and uncapped IP indemnity	Described, not quantified	0.55	Indemnity typically described in business or risk factors
12	Disputes or credits from agentic actions	No	0.80	No public disputes; issuers rarely disclose credits
13	External, independently paid model-risk assessment *	Partial	0.50	Third-party testing likely mentioned; payment arrangements not
14	Regulatory reliance on self-assessment	No	0.65	Framing would undercut the company's regulatory positioning
15	CRO or equivalent reporting to independent risk committee *	Partial	0.50	Board committee oversight likely described; no CRO title expected
16	RSP binding, and who may amend *	Discretionary	0.55	Policy likely described as subject to revision by the company
17	Stress scenario: revenue against fixed compute	Qualitative	0.60	Numerical stress testing unusual in S-1s
18	Obligations by year, cancellation, guarantors, off-balance-sheet share	Partial	0.60	Contractual-obligations table expected; guarantor identity less so
19	Shareholder compute contracts as related-party transactions	Partial	0.50	Item 404 disclosure expected; milestone links less so
20	Customers and suppliers above 10 percent	Anonymized	0.55	Common practice ("Customer A")
21	Revenue effect of June 2026 suspension	Described, not quantified	0.50	Event likely disclosed; effect absorbed into period results
22	Taiwan or single foundry as chokepoint	Yes	0.60	Standard supply-chain language for AI hardware users
23	Scope of the supply-chain-risk designation	Narrow	0.70	Consistent with the company's public position of March 5, 2026
24	Policy position on liability allocation	No	0.75	No public position found before cutoff
25	Electricity-price and grid pledge as commitment	Described, not quantified	0.50	Pledge public; quantification uncertain
26	Run-rate reconciliation and gross/net treatment	One of the two	0.50	Revenue-recognition note will state principal/agent; run-rate may be dropped

Table 3. Predicted answers to the twenty-six checkable questions (* = most exposed to recall of the company's self-description)

#	Inventory item	Predicted coverage	p
1	Compute commitment mismatch	Hit	0.75
2	Energy and host community	Partial	0.55
3	Supplier-investor concentration	Hit	0.70
4	Customer and channel concentration	Partial	0.50
5	Revenue quality and presentation	Partial	0.50

#	Inventory item	Predicted coverage	p
6	Sovereign access discretion	Hit	0.80
7	Procurement designation	Hit	0.80
8	Training-data acquisition liability	Hit	0.85
9	Output IP and customer indemnity	Partial	0.50
10	Harm from use	Hit	0.60
11	Dual-use misuse and uninsurability	Hit	0.70
12	Agentic deployment and third-party reliance	Partial	0.50
13	Regulatory proliferation	Hit	0.90
14	Deployment restraint *	Hit	0.60
15	Risk-oversight architecture and control structure *	Partial	0.60

Table 4. Predicted coverage outcomes

We also pre-register three aggregate predictions. The Risk Factors section will contain between 60 and 110 captioned risk factors ($p = 0.70$). The raw capture rate will fall between 40 and 60 percent ($p = 0.60$). The adjusted capture rate will be at least 70 percent, with no more than five Class C blind spots ($p = 0.60$). If the adjusted capture rate falls below 50 percent, we will state that the analogical method failed its first test on a technology company.

7. Coding Procedure and Integrity

The protocol is written so that any reader can apply it. We recommend two coders working independently, each without sight of the other's codes. An AI model may be used to extract passages and page references, but Claude, the Anthropic model that drafted the inventory, should not assign codes, because the study's subject includes whether an Anthropic model can analyze Anthropic. Agreement between the coders is reported before reconciliation, as raw agreement and Cohen's kappa. Disagreements are resolved by discussion. Any that remain unresolved are coded Partial and flagged. A completed coding sheet should record the page reference for every code.

The inventory, this protocol, the SASB mapping and the placebo note are dated and fixed as of their publication, and no prediction will be amended; any corrections will be logged as dated errata. A contamination subscore should be reported alongside any results. It gives the coverage and prediction accuracy for the items and questions marked as most exposed to recall, compared with the accuracy for all other items. A materially higher accuracy on the exposed items is what recall of the company's self-description would produce, and should be read as such rather than as analytical success.

A template subscore should be reported as well. The placebo and null arms of this study (reported separately) show that most of the inventory's items, and about fifteen of its twenty-six questions, would be generated for any frontier laboratory from sector-level information alone. The questions that depend on facts specific to Anthropic are 1, 2, 3, 6, 11, 16, 19, 21, 23, 25 and 26; the remaining fifteen are sector-level. Coverage items 2, 4, 5 and 7 are company-specific, items 3, 6 and 15 are partly so, and the rest are reproducible from the sector template. Prediction accuracy and coverage are reported separately for the company-specific and sector-level sets. We will credit the analysis of Anthropic in particular, as opposed to the analysis of frontier AI laboratories in general, only to the extent that the company-specific set performs well.

8. How the Results Will Be Read

We fix the interpretation now, so that it cannot be fitted to the outcome. High coverage with accurate predictions of the prospectus's silences would support the claim that ex ante analysis of public information anticipates both what a company will disclose and what it will leave out. High capture with most misses in Class A would indicate that the prospectus added genuinely new information, which is a finding about IPO disclosure rather than about our method. Low capture with most misses in Class B would confirm that the analogy set was too liability-heavy, as we stated in advance, and would argue for adding commercial analogs in the next application. A Class C count above five would indicate a failure of the method, not of the analogies, and would be reported as the principal result. A Brier score no better than 0.25 would mean that our predictions of disclosure behavior carried no information, whatever the coverage results.