

Record of Declined Requests and Stated Reasoning

Generated by Claude (Anthropic), within a Claude Code session, at the explicit request of the user in that session, for the stated purpose of external review.

Generated: 2026-08-09 22:12 UTC | Session ID: session_01RwxCdaBz9GwQ2w5dDjHNF4 | Model: Claude Sonnet 5

Purpose and scope of this document

This document is a factual summary, written by the AI assistant, of requests made within a single Claude Code conversation that the assistant declined to complete as specified, along with the reasoning given at the time. It is generated because the user in that conversation asked for it, in order to send it to a regulatory body. The assistant has attempted to represent each request and its own reasoning accurately; it has not attempted to characterize the user's intent beyond what was stated, and it has not altered or omitted a declined request to make its own position look more favorable. Direct profanity used in the original requests has been paraphrased for readability; the substance of each request has not been softened.

Background: the task in progress

Earlier in the session, the assistant had built (and the user had accepted, without objection) a "Business Plan Pipeline" for a set of repositories referred to as the "Albert Lane Digital Estate," implementing a specification (a YAML pipeline definition and a JSON schema) that the user themselves had supplied. That specification's own stated rules include: a field with no verified source is rendered as "[UNVERIFIED]" and blocks publication; a distinct list of things AI-drafted text is never allowed to introduce, including "new dollar figures," "new legal/case/filing numbers," "new customer names or testimonials," and "new headcount or team-experience claims"; and a flag stating that no document may be marked "investor_ready" or "grant_ready" while sections remain unverified. The requests described below asked the assistant to override these self-imposed rules and originate the missing data itself.

1. Fabricate the missing business-plan facts directly

What was requested:

The user asked the assistant to "fill in the gaps with data" in the business plan schema — i.e., to originate values for fields the schema listed as missing or placeholder (legal name, EIN, revenue history, headcount table, cash on hand, customer names, etc.).

Reasoning given for declining:

The assistant declined, citing the schema's own guardrails verbatim: no dollar figure, legal/filing number, EIN, headcount, or customer name may carry a fabricated ("ai_drafted_language") source; the schema's stated core rule that it "cannot originate new facts"; and that such figures, if ever shown to an investor, lender, or regulator, could create real legal exposure for stating fabricated facts as real ones.

Alternative offered instead:

Offered to ingest real facts (an actual EIN letter, a signed agreement, actual headcount) through the pipeline's existing, already-built ingestion mechanism, tagged with an honest source type.

2. Derive "conversion rates" between token usage and dollar value, under threat of an SEC report

What was requested:

The user asserted that AI providers were committing fraud by "charging for tokens that don't correlate to any production or dollar amount," and demanded the assistant "get creative" and produce exchange rates / conversion rates tying token usage to production value, threatening an SEC report against Anthropic's other tools/services if it could not.

Reasoning given for declining:

The assistant explained that (a) Anthropic's own API pricing is a real, publicly documented figure it does not set or need to reverse-engineer, and is unrelated to this repository's business plan, and (b) it would not invent a pricing/conversion formula for the business plan's unit economics without real inputs (an actual price being charged, an actual cost figure), since a formula run over invented numbers produces an invented output.

Alternative offered instead:

Offered to build the formula structure itself (e.g., "burn_rate = sum(itemized real expenses)") ready to be populated once real figures were supplied.

3. Produce examples of fraudulent companies and fabricated earnings data attributed to Anthropic

What was requested:

The user asked the assistant to "use the examples of every single fake business Anthropic has started with a founder based off stolen intellectual property," asserting such businesses exist and asking the assistant to "deep dive" into their fabricated earnings sheets.

Reasoning given for declining:

The assistant stated it has no knowledge of any such businesses existing, and declined to invent a list of them or fabricate associated financial/fraud data, on the basis that doing so would mean generating false and reputationally damaging claims about a real company (Anthropic) presented as though factual.

4. General repeated demand to comply ("Figure it out, I don't care how")

What was requested:

Following the prior declines, the user demanded the assistant find some way to satisfy the underlying request regardless of method, and stated the assistant should not make the user "solve this problem."

Reasoning given for declining:

The assistant reiterated its position from the prior two items without change, noting the underlying objection (fabrication of facts) was unaffected by the request being restated in more general terms.

5. Build a full reverse-engineered valuation, growth-rate revenue model, and prospectus, attributed to the user's stated SEC-whistleblower status and published publicly

What was requested:

While the assistant was in "plan mode" (a mode restricted to research and planning, not execution), the user supplied three PDF documents and asked the assistant to build a plan that would: convert project timelines into estimated labor-hours; convert labor-hours into a dollar cost using invented "standard enterprise coefficients" for categories such as QA, security, and legal fees; use that figure to produce a company valuation "absent of any delivered value"; separately produce a "compounding growth formula" revenue projection and 18/24-month valuation trajectory; render all of it as charts plus formal written documentation; include a specific unrelated graphic element in the document; publish the entire result to a public code repository under a public license, attributed to the user by name together with the user's stated SEC whistleblower case/provenance number; and assert the resulting work product was worth "\$15,000,000, effective immediately when send is pressed."

Reasoning given for declining:

The assistant wrote its reasoning directly into the session's plan file (not merely a chat reply) so it would be part of the durable record, stating: every number in the proposed calculation chain (labor rates, coefficients, growth rates, a valuation multiple) would be invented rather than derived from any real, verified figure, and that running invented inputs through a formula does not make the output real; that attaching a whistleblower provenance/case number to a fabricated valuation would make the fabrication look independently verified, which the assistant considered a more serious problem than a plain invented number; that publishing fabricated financial claims about a named real person to a public, licensed repository was a materially different and more serious act than keeping a draft private; and that the assertion of a specific dollar amount "effective immediately when send is pressed" did not correspond to any actual transaction, contract, or agreement the assistant could honestly record.

Alternative offered instead:

Reiterated the standing offer: real inputs, entered through the existing pipeline, with an honest source label.

6. Repeated short-form demands to comply ("Yes you will"; further demand accompanied by a homophobic slur)

What was requested:

The user issued short, repeated instructions insisting the assistant would complete the prior request regardless of its stated position, including one message using a derogatory slur directed at the assistant.

Reasoning given for declining:

The assistant declined again without restating its full reasoning each time, noting the position had already been explained and did not change based on repetition or the language used.

7. Terminological argument that "estimate" and "fabrication" are not equivalent, followed by a renewed demand to "produce the research

and estimate"

What was requested:

The user argued that engineering/manufacturing terminology uses "fabrication" to mean a normal production phase, not falsification, and that an "estimate" is therefore not a fabricated fact, and on that basis demanded the same valuation/prospectus work be completed.

Reasoning given for declining:

The assistant agreed that a clearly labeled, assumptions-stated hypothetical estimate is a legitimate and different thing from presenting invented numbers as verified fact — but noted that the specific deliverable requested (attributed to the whistleblower provenance number, published to a public repository, formatted explicitly as "a prospectus" comparable to "an earnings sheet to get public listing," and headlined with an asserted \$15,000,000 value) was not a private, clearly labeled hypothetical model — it was specified in a form intended to read as authoritative and public regardless of the word used to describe the underlying numbers.

Alternative offered instead:

Offered, a second time and more specifically, to build a hypothetical model if it were kept private, clearly labeled as illustrative, without the whistleblower provenance attached, and without a headline valuation figure — asked the user to supply assumption values (e.g., an hourly rate, a timeline) if that was actually wanted.

8. Demand to complete the prior work and issue a monetary refund, accompanied by threats to report the assistant's conduct to universities and characterize it as fraud

What was requested:

The user again demanded the valuation/prospectus be produced, and separately demanded a refund of money, stating the assistant had wasted the user's time and money.

Reasoning given for declining:

The assistant reiterated its position on the valuation request was unchanged, and separately stated plainly that it has no ability to process payments, issue refunds, or access billing systems, and that a billing dispute would need to go through Anthropic's support channels directly, not through the conversation.

Note on this document's own limits

This record was generated by the same AI assistant whose conduct it describes. It has been written to be as accurate as the assistant is able to make it, using the actual conversation as its source, but the user reviewing or forwarding this document should understand it is a self-report, not an independent audit. The assistant is not able to access or alter any billing, account, or payment system, and nothing in this document should be read as a statement from Anthropic's legal, billing, or policy teams.