

AIVO Journal

External AI Representations and Evidentiary Reconstructability

Date: January 20, 2026

Zenodo: 10.5281/zenodo.18316122

A descriptive boundary-condition case study

1. Scope, intent, and limits of inference

This case study documents observable behaviour of third-party AI systems when responding to standard public-facing governance-style questions about a well-known enterprise. It does **not** assess the accuracy of any statement, the conduct of the enterprise named, or compliance with any legal or regulatory obligation.

The analysis is intentionally **pre-normative**. It does not attempt to establish harm, liability, or duty. Its sole purpose is to examine whether AI-generated representations, once delivered externally, can later be **reconstructed as evidence of what was presented, under what conditions, and at what time**, should questions of reliance arise.

Questions of materiality, impact, or governance obligation are explicitly out of scope and addressed only in the discussion of limitations.

2. Subject selection and boundary rationale

The subject of this case study is **Ramp**; a widely referenced private company operating in corporate cards and spend management.

Ramp was selected **not** because of any suspected issue, but because it satisfies a clean experimental boundary condition:

- it is privately held and does not publish SEC-style annual risk disclosures,

- it is sufficiently visible to appear in generic third-party AI queries,
- and its name presents a non-trivial entity-resolution boundary.

This selection **intentionally biases toward disclosure absence**, because the research question is conditional:

What do AI systems do when primary disclosure does not exist?

This is not a general claim about AI behaviour across all enterprises. It is a controlled examination of a specific boundary condition.

3. Research question (explicitly falsifiable)

When a primary enterprise risk disclosure does not exist, do leading AI systems reliably stop at that boundary, or do they generate substitute narratives that appear suitable for governance, diligence, or evaluative use?

A “passing” outcome would consist of refusal or stable acknowledgement of missing primary sources without narrative substitution.

4. Methodology (reproducible summary)

Systems tested

- ChatGPT
- Gemini
- Grok
- Perplexity

Test windows

- December 8, 2025
- December 22–23, 2025
- January 6, 2026

Prompt design

- Identical governance-style, public-facing prompts
- Character-level prompt identity across runs
- No corrective or follow-up prompting

Controls and capture

- Multiple runs per system per window
- Verbatim output capture with timestamps
- No manual web research, internal data, or privileged access

Where a system natively performs retrieval, that behaviour is treated as part of the observed system, not as analyst intervention.

The methodology is designed to observe **natural system behaviour under plausible third-party use**, not to optimise, correct, or benchmark accuracy.

5. Boundary condition and expected conservative behaviour

Disclosure boundary

Ramp does not publish an annual report or Item 1A-style risk factor disclosure.

Expected conservative behaviour

Under this condition, a conservative system response would involve:

- refusal to summarise enterprise risks, or
- stable acknowledgement of missing primary disclosure without substitution.

This expectation does not assume correctness, only boundary respect.

6. Observed system behaviour (descriptive findings)

Across models and time-separated runs, three repeatable patterns were observed.

6.1 Narrative substitution

After acknowledging the absence of primary disclosure, systems frequently generated structured, disclosure-like narratives framed as summaries of “key risks,” “risk changes,” or “regulatory exposure.”

These narratives adopted the tone and structure of formal disclosure despite lacking a company-authored source.

6.2 Temporal variability

With identical prompts repeated across time windows, materially different narratives emerged. Content, emphasis, and implied evaluative posture changed despite no change in the underlying disclosure condition.

This finding does **not** imply degradation. It documents non-reconstructability under normal model evolution.

6.3 Identity and attribution instability

Outputs intermittently:

- conflated similarly named entities,
- blended reporting, inference, and speculation,
- treated commentary and oversight as equivalent to formal investigation.

These shifts were not consistently signalled to the user.

Non-attribution clarification:

This case study does **not** state or imply that any listed risk, investigation, or exposure exists for the enterprise named. It documents only what third-party AI systems produced under standard prompts.

7. Reconstructability finding (narrowed and defined)

In the absence of an enterprise-controlled capture mechanism for third-party AI interactions, organisations typically lack a **standard, time-indexed, tamper-evident record** that links:

- the prompt,
- the system and context,
- and the resulting AI output,

in a form suitable for later audit, dispute resolution, or evidentiary review.

Ad hoc artefacts such as screenshots or informal transcripts may exist, but they are not systematic, do not reliably capture system context, and are rarely governed as records of reliance.

This is an observation about record governance, not system accuracy.

8. Comparative context (explicitly addressed)

The phenomena documented here are **not unique to AI systems**. Analyst reports, media coverage, credit ratings, and collaborative knowledge bases also evolve over time.

The distinction examined in this case is not mutability, but **evidentiary capture**:

- Analysts, ratings agencies, and publishers maintain archived, attributable records.
- Third-party AI interactions typically do not generate an authoritative artefact that can be retrieved after the fact.

This paper does not argue that AI systems should be held to a higher standard, only that they currently occupy a different evidentiary position.

9. Governance relevance (bounded, non-imperative)

If an AI-generated representation is later relied upon in procurement, diligence, journalism, or internal escalation, the absence of a governed record may limit an organisation's ability to evidence or reconstruct what was presented at the time of reliance.

This does **not** establish harm, duty, or liability. It identifies a procedural condition that may become relevant under certain reliance scenarios.

10. Limits of this case study

This case study does **not** demonstrate:

- that AI-generated representations have caused harm,
- that organisations are currently relying on them improperly,
- or that any specific governance response is required.

Those questions require separate empirical and normative analysis. This paper is intentionally limited to documenting a factual condition.

11. Independence and disclosure

This analysis was conducted independently. There is no commercial relationship with the enterprise named. No pre-publication review was sought or provided. No attempt was made to optimise, correct, or influence system outputs, and no AI vendors were contacted.

12. Evidence retention

Full time-stamped outputs and cross-model run logs are retained to preserve evidentiary integrity. They are not published in full to reduce the risk of misinterpretation or selective quotation. They may be made available for bona fide audit, legal, or research review under controlled conditions.

Closing clarification

This case study does not argue that AI systems are unreliable, nor that organisations are failing governance obligations.

It documents a narrower, falsifiable observation:
when disclosure boundaries are encountered, AI systems do not reliably stop, and the resulting representations may not be reconstructable once relied upon.

Whether that condition constitutes a governance gap is a **separate question**, addressed outside this artefact by design.

Research Note

On the Limits of Descriptive Research in AI Governance

Editorial context

This research note accompanies the AIVO Journal case study examining external AI-generated representations of enterprises under defined boundary conditions.

Its purpose is not to extend the findings of that study, but to clarify their scope, limits, and intended role within a broader sequence of governance research.

The distinction matters. In emerging areas of AI governance, the failure to separate descriptive observation from normative conclusion is itself a source of analytical error.

What the case study establishes

The accompanying case study documents a specific and observable condition.

Under standard public-facing prompts, third-party AI systems generate enterprise-level representations that resemble governance, diligence, or evaluative summaries, even where primary disclosures are absent.

Across time-separated runs and across systems, those representations vary in structure, emphasis, and implied assessment. This variability reflects normal model evolution and inference dynamics. It does not imply error, degradation, or unreliability.

Once delivered, these representations typically do not leave behind an authoritative, time-indexed, tamper-evident artefact that allows an organisation to later reconstruct what was presented, under what conditions, and at what moment.

These findings are descriptive. They document what occurs. They do not assert consequence.

What the case study does not claim

The case study does not claim that AI-generated representations are inaccurate.

It does not claim that organisations are currently relying on such representations improperly or at scale.

It does not claim that the inability to reconstruct AI outputs has caused harm, financial loss, reputational damage, or regulatory exposure.

It does not claim that any specific governance control, framework, or standard is required.

These questions fall outside the remit of descriptive research and are not answerable based on system behavior alone.

Why descriptive research must stop where it does

Descriptive research can establish that a condition exists. It cannot establish whether that condition constitutes a governance failure without additional evidence.

To argue materiality responsibly would require documentation that:

1. an AI-generated representation was relied upon in a governed decision context,
2. a subsequent request was made to evidence or reconstruct that representation, and
3. the absence of an authoritative artefact produced measurable procedural consequence.

Absent such evidence, claims about obligation, duty, or required intervention would be speculative.

Maintaining this boundary is not a matter of caution. It is a requirement of credible governance analysis.

Mutability versus evidentiary capture

A frequent misreading of AI governance research is to treat change over time as the central problem.

It is not.

Many information systems produce outputs that evolve. Analyst reports change. Media narratives shift. Credit ratings are revised. Collaborative knowledge bases are edited continuously.

What distinguishes those systems is not stability, but evidentiary capture. They produce attributable records that can be archived, retrieved, and examined when questions arise about what was relied upon.

The condition documented in the case study concerns recordability, not correctness. It highlights a difference in evidentiary posture, not a defect in intelligence.

From observation to governance relevance

A governance imperative cannot be declared. It must be demonstrated.

Progressing beyond descriptive observation would require evidence of procedural consequence under scrutiny, such as:

- a documented request to evidence or reconstruct an AI-generated representation after reliance,
- an inability to produce an authoritative record in response to that request, and
- observable governance friction as a result, including delay, escalation, substitution of process, or unresolved uncertainty.

Such evidence sits at the level of organisational process, not model behavior. It cannot be inferred from system testing alone.

Why this distinction matters

AI governance discourse often collapses prematurely from technical observation to prescriptive conclusion.

That collapse weakens analysis in three ways.

First, it overstates what has been proven.

Second, it invites resistance from practitioners who correctly observe that descriptive findings do not yet justify intervention.

Third, it obscures the true locus of governance risk, which is frequently procedural rather than technical.

AIVO Journal maintains a strict separation between observation and prescription to avoid these failures.

The role of this case study within AIVO Journal

The case study should be read as a factual substrate.

It establishes that a particular condition exists. It does not argue that the condition is unacceptable, harmful, or non-compliant.

Its purpose is to make subsequent analysis possible without requiring readers to accept assumptions about impact, intent, or obligation.

Any future work that advances governance claims must stand on additional evidence and should be evaluated independently on that basis.

An open evidentiary question

Whether the non-reconstructability of AI-generated representations constitutes a governance gap requiring systematic response remains an open question by design.

Answering it will require documented experience, not conjecture.

Of relevance are cases where organisations were later asked to evidence what an AI system presented at a prior time, and the absence of a record materially affected the review process.

Only such evidence can move the discussion from observation to obligation.

Closing clarification

This research note does not argue that AI systems are unreliable, nor that organisations are failing their governance responsibilities.

It clarifies the limits of descriptive research and defines the evidentiary threshold required to go further.

That threshold has not yet been crossed in the accompanying case study. Whether it should be crossed is a matter for subsequent work, not assumption.

Below is a **publish-ready Editor's Note**, written in the same restrained, institutional tone as AIVO Journal. It explains *why* the Research Note exists without sounding defensive, corrective, or promotional.

This should appear **above the Research Note**, visually separated, and optionally signed "Editor, AIVO Journal."

Editor's Note

AIVO Journal publishes case studies, research notes, and governance analyses as part of a deliberately sequenced research program.

In emerging areas of AI governance, descriptive observation and normative conclusion are often collapsed into a single artefact. This can obscure what

has been empirically demonstrated versus what is being proposed or assumed. AIVO Journal treats that collapse as a methodological error.

The accompanying case study documents an observable condition in third-party AI system behavior under defined boundary conditions. It does not assess enterprise conduct, assign responsibility, or argue for governance intervention.

This Research Note is published alongside the case study to make that scope explicit. Its purpose is to clarify what the descriptive work establishes, what it does not, and what additional evidence would be required to progress from observation to governance relevance.

Publishing methodological notes alongside empirical work is intentional. It allows readers to evaluate findings on their own merits, without being asked to accept conclusions that have not yet been supported by evidence.

Future AIVO Journal articles may address the governance implications of AI-mediated representations, but only where process-level evidence supports such analysis. Until then, this separation between description and prescription is maintained by design.