

**From AI Answers to Auditable Decisions:
An Evidence Framework for Business Analysis**

Sruthi Rachamalla

Southern Illinois University, Carbondale

ORCID iD: 0000-0001-9312-507X

Abstract

Background: A business recommendation can sound convincing while obscuring the evidence, assumptions, and disagreements behind it. Citations help readers locate information, but they do not establish whether a proposed decision follows from that information.

Methods: This paper develops a conceptual evidence framework for business analysis. It separates source observations, interpretations, assumptions, recommendations, and accountable decisions. Versioned records connect each material claim to supporting passages, conflicting evidence, and the person responsible for review.

Results: The proposed architecture combines a controlled evidence registry, claim assessment, and a decision record. A fictional product prioritization example demonstrates how limited interview evidence can support a bounded pilot without supporting a general rollout. The evaluation plan tests evidence accuracy, omitted objections, reviewer effort, and the ability to reconstruct decisions.

Conclusion: Auditability requires more than attaching references to an answer. Organizations need to preserve how evidence was interpreted, which uncertainties remained, and who accepted them. The framework is a design proposal; no deployment results are reported.

Keywords: Business analysis; decision traceability; evidence provenance; human oversight; enterprise AI

1 Introduction

A product team receives a polished recommendation to introduce bulk data export. The summary quotes several customers and concludes that the feature should take priority. Before committing resources, the team needs to know whose views were represented, which objections were excluded, and whether the proposed benefit was measured or assumed.

Citation quality and factual correctness are distinct evaluation concerns [1]. A reference can be genuine yet fail to support the conclusion attached to it. NIST likewise treats accountability and transparency as governance concerns, not merely presentation features [2]. This paper proposes an evidence trail that lets reviewers inspect the transition from source material to organizational commitment.

2 Methods

The unit of analysis is a material claim: a statement that could change the decision if it were wrong. Each claim receives an identifier, a type, a scope, and an evidence status: supported, contested, or unverified. Separating compound statements makes individual support checks possible, consistent with atomic factual evaluation [3]. Status describes the reviewed evidence, not model confidence.

Five types are kept distinct: source observations, interpretations, assumptions, recommendations, and decisions. Recommendations depend on objectives and tradeoffs; they are not facts awaiting a citation. The framework links sources, transformations, and responsible actors using provenance concepts [4]. It is specified through record requirements and a fictional workflow, rather than an experiment. Its proposed contribution is making decision dependencies inspectable without treating traceability as proof of correctness.

3 Results

3.1 Evidence architecture

Figure 1 organizes the workflow around three components. An evidence registry stores authorized source versions, dates, passage locations, and collection context. A claim service links statements to supporting and conflicting passages. A decision workspace records alternatives, unresolved assumptions, reviewer judgments, and the selected course of action.

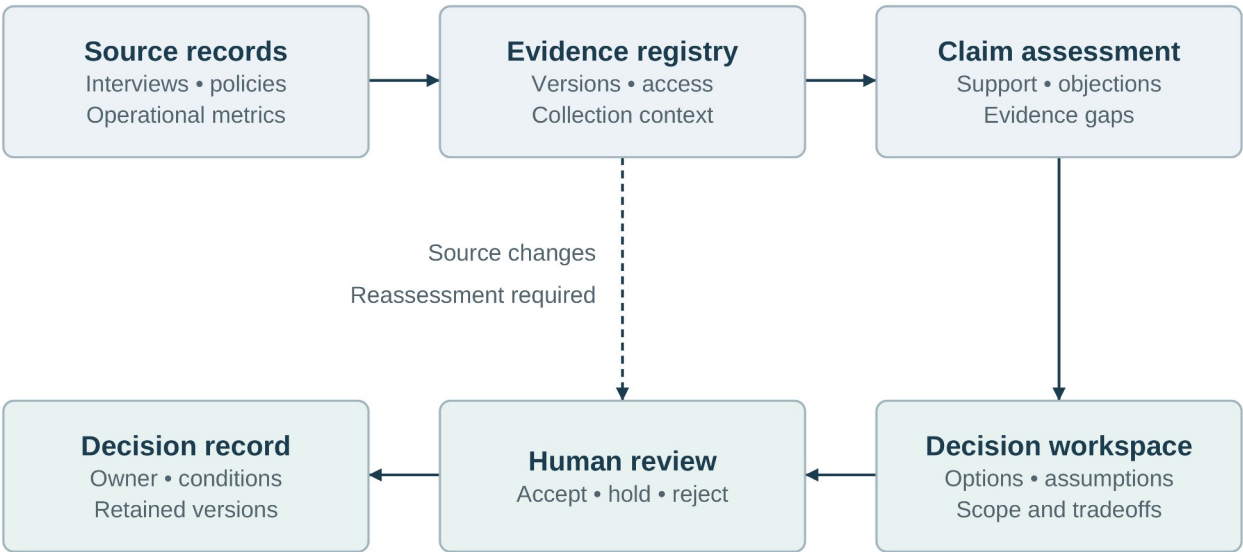


Figure 1. Evidence-to-decision architecture. Source changes trigger human reassessment while earlier records remain preserved.

Each record includes the business question, intended audience, evidence identifiers, claim types, alternatives, and decision owner. A reviewer can move from a recommendation to a particular source passage without searching an entire document. Supporting and opposing passages receive equally visible links; unresolved disagreement remains part of the record.

3.2 Evidence checks and human review

The source set and retrieval scope are recorded before synthesis. Reviewers search for objections within that scope and document missing stakeholder groups. Repeated summaries of one

interview count as one underlying source, not independent corroboration. Restricted evidence retains its access controls; summaries must not expose information their readers cannot access.

Automated checks can verify that a passage exists, matches the recorded version, and supports a quotation. Whether it justifies a broader interpretation requires assessment. A source can faithfully record a mistaken belief. Reliability, relevance, and independence therefore remain separate from accurate transcription.

Before accepting a recommendation, a reviewer examines its decisive claims, strongest objection, and unresolved assumptions. In a controlled study, deliberative interventions reduced overreliance, with usability tradeoffs [5]. Here, the proposed interface asks reviewers to record why they accepted the recommendation and what evidence would change their decision.

Approval binds the decision to its evidence snapshot, scope, owner, and review date. Missing critical evidence produces a hold or a narrower proposal, not a higher confidence score. Source corrections flag dependent decisions for reassessment; earlier records remain distinguishable from current ones. Retention and redaction policies govern preserved excerpts. The record preserves the reviewer's judgment rather than merely repeating the recommendation.

3.3 Worked example

Consider twelve fictional customer interviews. Seven participants favor bulk export without further conditions, three support it only with additional access controls, and two prefer improved reporting. Five of the seven unconditional supporters are large-account administrators recruited by sales. The pattern identifies a need to investigate; it does not establish demand across the customer population. Table 1 distinguishes the resulting statements.

Table 1. Distinct record types in the fictional export decision.

Record / ID	Example	Review requirement
Observation / O-01	Seven support export outright.	Verify interview passages and sample size.
Interpretation / I-01	Need may vary by role.	Check recruitment and objections.
Assumption / A-01	Export saves reporting time.	Measure the proposed benefit.
Recommendation / R-01	Run a restricted pilot.	Compare alternatives and risks.
Decision / D-01	Conditional pilot approved.	Name owner and review date.

Decision D-01 approves recommendation R-01, linked to observation O-01 and assumption A-01. In this fictional record, O-01 links to interviews INT01–INT07, paragraph 4 of each record. The reviewer chooses a restricted pilot rather than a general release. Permission testing is a prerequisite; adoption and task completion time are follow-up measures. No time-saving claim is accepted before measurement. The record also preserves improved reporting as an alternative, with a reason for deferring it. Different business priorities could reasonably produce another decision from the same evidence.

4 Discussion

Evaluation should compare ordinary cited summaries with structured evidence records using identical source packets and model settings. Independent analysts establish reference claims, objections, and acceptance criteria before testing. Cases include misleading quotations, duplicated sources, omitted objections, unsupported forecasts, and later corrections. Comparable reviewer groups assess cases in counterbalanced order.

Report the proportion of factual claims supported by evidence, known objection coverage, and whether recommendations disclose assumptions and tradeoffs. Count erroneous approvals, unnecessary holds, and reviewer time. An approval is erroneous when it breaches a predefined factual, policy, or evidentiary requirement—not merely when another reviewer prefers a different option. Include unapproved tasks and separate error severity. A second reviewer should reconstruct the decision without its author. Model agreement cannot replace independent evidence assessment.

The framework adds storage and review effort. It cannot recover missing interviews, guarantee honest sources, or settle competing business values. A complete trail can document a poor choice. Begin with one consequential workflow and verify critical dependencies manually. Auditability supports challenge and correction; it does not certify a decision.

5 Conclusion

Useful business analysis preserves the boundary between what was observed, what was inferred, and what was chosen. An evidence-linked decision record makes that boundary examinable. Its practical value is not a more persuasive answer, but a decision that another qualified person can question, reconstruct, and reconsider.

References

- [1] T. Gao, H. Yen, J. Yu, and D. Chen, “Enabling large language models to generate text with citations,” in Proc. EMNLP, 2023, pp. 6465–6488. doi: 10.18653/v1/2023.emnlp-main.398.
- [2] E. Tabassi, “Artificial Intelligence Risk Management Framework (AI RMF 1.0),” NIST AI 100-1, 2023. doi: 10.6028/NIST.AI.100-1.
- [3] S. Min et al., “FActScore: Fine-grained atomic evaluation of factual precision in long form text generation,” in Proc. EMNLP, 2023, pp. 12076–12100. doi: 10.18653/v1/2023.emnlp-main.741.
- [4] L. Moreau and P. Missier, Eds., “PROV-DM: The PROV data model,” W3C Recommendation, Apr. 30, 2013. [Online]. Available: <https://www.w3.org/TR/prov-dm/>
- [5] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, “To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making,” Proc. ACM Hum.-Comput. Interact., vol. 5, no. CSCW1, Art. no. 188, pp. 1–21, 2021. doi: 10.1145/3449287.