

Correction: Retention, Metadata Privacy, and Scope of the Embedding-Financed Proxy Company Model

Corrections to the published note and a narrower statement of its contribution

Anonymous author

25 October 2026

Abstract

We correct response-dependent control, private transcripts, and joint metadata leakage in the proxy-company note. We give a response-marginalized retention criterion, specify observation boundaries, and reclassify standard facts and assumed revenue. The higher categorical framework is an interpretation, not evidence of privacy or profit.

Status and scope of this correction

This corrigendum concerns Anonymous author, *Higher Categorical Semantics of an Embedding-Financed Proxy Company*, Higher Categorical Semantics, 25 October 2026, whose published label is **Editorially reviewed**, not peer reviewed:

<https://higher-categorical-semantics.pages.dev/research/higher-categorical-semantics-of-an-embedding-financed-proxy-company-9083e039>.

References to sections and result numbers below refer to that note [O]. This correction supersedes its implementation-level application of Proposition 3.2, its underspecified transcript and independence claims, and any interpretation that its standard lemmas or consistency example establish new foundational theory or commercial viability. The original displayed arithmetic and the elementary factorization, decoding, privacy-composition, and reserve identities remain valid under their stated mathematical hypotheses.

The original framing overstated the evidence supplied by the formalism. A specification is not its implementation, a declared metadata exception is not embedding-only retention, and an assumed buyer payment is not a derivation of demand. This correction supplies explicit obligations and counterexamples, rather than claiming these obligations have been met by a production service.

1. Replace the application of Proposition 3.2

The implication in the original proposition is valid: if all relevant selections and updates depend only on the listed permitted inputs, the induction works. Its hypothesis, however, was not established for the advertised protocol. A response-dependent retry, charge, log, or crash path need not satisfy it. A counterexample violating the hypothesis does not refute the implication; it refutes an unconditional application to that implementation.

Merely adding the full response to the permitted representation changes the content-retention contract. Restricting to successful requests also does not suffice if successful paths retain response-dependent

information. We instead replace the implementation claim with the following explicit finite criterion.

Let A be the finite set of full durable states, X encoded requests, Z a **fixed, independently specified** permitted representation, and $q : X \rightarrow Z$ its map. Fix public exogenous inputs throughout. Let Y contain the entire bounded upstream response/retry transcript, including terminal failures. Suppose

$$K(y \mid a, x), \quad U(a' \mid a, x, y)$$

are normalized kernels for the upstream transcript and the resulting durable update. All persistent billing, counters, indexes, retained parameters, and recovery records belong to a' . Raw responses may influence volatile execution; their durable effect is assessed through

$$L(a' \mid a, x) = \sum_{y \in Y} K(y \mid a, x) U(a' \mid a, x, y). \quad (1)$$

Replacement lemma (response-aware retention). A normalized kernel $\bar{L}(a' \mid a, z)$ on $z \in \text{im } q$ satisfies $L(a' \mid a, x) = \bar{L}(a' \mid a, q(x))$ exactly when

$$q(x) = q(x') \implies L(a' \mid a, x) = L(a' \mid a, x') \quad \text{for all } a, a'. \quad (2)$$

Proof. Necessity is substitution. For sufficiency define $\bar{L}(\cdot \mid a, z)$ using any representative x of z . Condition (2) makes it independent of the representative; normalization follows from (1). It is unique on the image. If (2) holds at every request boundary, volatile state is reset at that boundary, the initial durable law depends only on fixed public inputs, and subsequent selection of such kernels uses only durable state and permitted inputs, induction gives factorization of durable path laws through the ordered permitted-input sequence. Include all bounded failure outcomes to obtain a total kernel. $\square$

This is an application of the standard quotient-factorization criterion, not a novel foundational theorem. It permits response-dependent internal control only when the **aggregated durable law** respects the specified quotient. It makes no retention promise about an upstream provider or an observer of excluded transient network traffic.

Retry counterexample. Take $x = (p, s)$, $q(p, s) = p$, response $y = p \oplus s$, and retain a counter $r = y$ indicating a retry. Requests $(0, 0)$ and $(0, 1)$ have the same permitted representation but yield $r = 0$ and $r = 1$. Thus (2) fails. Erasing the response bytes does not erase the counter's information. If actual upstream payments change the durable cash balance, that balance must pass the same test or be an explicitly stated exception to the stronger contract.

A permitted metadata policy must therefore be fixed before checking (2). Enlarging q with whichever logs or metadata an implementation happens to retain is not a verification of embedding-only retention.

2. Replace the release-transcript specification in section 5

The instruction to include visible refusals and charges is correct as a requirement to analyze the whole observation. It was not sufficient to establish that the resulting transcript is private. The following condition must accompany it.

For a fixed analyst interface and declared user-level adjacency $D \sim D'$, write its full observable transcript as

$$\mathcal{T}(D) = (C(D), F(D), M(D)),$$

where C denotes charges, F refusal/authorization flags, and M released answers. If a deterministic C differs between adjacent datasets, the event that its value equals $C(D)$ has probabilities 1 and 0. No finite pure ε can satisfy the differential-privacy inequality. The identical argument applies to a deterministic private-data-dependent refusal. The original Proposition 5.2 correctly exposes this failure; a private M cannot repair a nonprivate companion field.

A valid replacement must adopt one of these explicit designs:

1. Charges and flags depend only on fixed public query parameters and are constant across adjacent datasets.
2. They are postprocessing of an already private transcript and data-independent randomness, without additional access to the protected dataset.
3. The complete joint mechanism, including charges and flags, is proved private with its own budget or an adaptive conditional composition argument.

Checking each marginal separately is insufficient. Nor does unspecified noise establish privacy: its support, sensitivity, randomness, and composition must satisfy the joint guarantee. These are the standard postprocessing and composition requirements [D].

Token charges for a fixed public request can be unchanged by adjacency on a separate private analytics dataset. They then create no additional leakage about that dataset. Charges computed from a private request or private dataset are a different observation model. A claimed guarantee must specify whose data vary, what is held fixed, and who sees the bill. Internal unnoised vectors and raw upstream traffic remain outside a central private-release guarantee; they are not thereby protected from their respective observers.

In the corrected finite example below, the analytics interface has a public fixed price and public fixed permission policy. This is a modeling choice, not a claim about token-priced production services.

3. Replace the independence assertion in section 8

The original conditions “ P is independent of the secret S ” and “metadata are independent of S ” are not sufficient for their **joint** transcript to hide S . This is a genuine gap in the stated independence argument.

Counterexample. Let P, S be independent fair bits and retain metadata $W = P \oplus S$. Then

$$P \perp S, \quad W \perp S, \quad S = P \oplus W. \quad (3)$$

Thus each field individually hides S , while their pair reconstructs it. In this example W is exactly the original model’s upstream response. Merely naming that response-derived bit metadata does not make the retention policy safe. Any further analytics output appended to (P, W) cannot remove this reconstruction attack.

Corrected sufficient condition. Let V denote the entire durable analytics transcript, including metadata, private releases, prices, and visible flags. Require

$$P \perp S, \quad \Pr(V = v \mid P = p, S = s) = Q(v \mid p) \quad (4)$$

on the supported inputs. The second condition is the Markov property $S \rightarrow P \rightarrow V$, not separate marginal independence. It is satisfied when all these fields are generated from P , fixed public inputs,

and fresh randomness jointly independent of (P, S) , with no unaccounted response-dependent durable channel.

Proof of joint non-disclosure under (4). Factor the joint law as

$$\Pr(S = s, P = p, V = v) = \Pr(S = s) \Pr(P = p) Q(v | p).$$

This is $\Pr(S = s) \Pr(P = p, V = v)$, proving $S \perp (P, V)$. $\square$

For the corrected worked example, retain P and independently generated public-policy records, erase the response $P \oplus S$, and release randomized response computed from P with fresh independent coins. Use the fixed public fee described above. Condition (4) now follows by construction. This non-disclosure statement concerns secret S ; randomized-response privacy concerns variation of P at the external release interface. The user receiving the service response and the upstream provider have different observations. There is no corresponding assertion for arbitrary embeddings, correlated priors, or response-dependent accounting.

4. Reclassify Theorem 6.1 and expose the actual descent obligation

If $P \hookrightarrow \mathbf{Cfg}$ is already a subobject in $\mathbf{Sh}_\infty(\mathcal{A}, J)$, its local-to-global property is standard sheaf descent. The original proof is valid but cannot supply new evidence that a concrete auditing policy is local. Reclassify the result as a standard descent lemma.

The substantive implementation question begins earlier: a candidate predicate may only define a **subpresheaf** P_0 of a configuration sheaf. One must prove that locally admissible matching configurations glue to a configuration satisfying the intended global predicate. Restriction stability alone is insufficient.

Nonlocal-budget example. On a two-point discrete site, let $\mathbf{Cfg}(U)$ be nonnegative resource allocations to points of U . Define

$$P_0(U) = \left\{ a : U \rightarrow \mathbb{R}_{\geq 0} : \sum_{i \in U} a_i \leq 1 \right\}.$$

This is a subpresheaf: restrictions preserve its bound. On the cover $\{\{A\}, \{B\}\}$ of $\{A, B\}$, local allocations $a_A = a_B = 3/4$ are admissible and agree on the empty overlap. Their unique glued configuration has total $3/2 > 1$ and is not admissible. Hence P_0 is not a sheaf. Sheafifying it changes the specified global budget; it does not enforce that budget.

Consequently, the correction is not to assert without proof that “the metadata policy is a subsheaf.” A metadata map and a predicate subobject have different types. One must define the relevant admissibility predicate, prove descent where it holds, and retain genuinely global constraints where it does not. A complete inventory of real storage surfaces is a separate empirical obligation.

5. Correct the categorical and homotopical scope

The complete Segal object, organizational Cartesian fibration, and optional BC_2 factor in section 2 did not establish the finite retention, decoding, privacy, or cash-flow conclusions. Appending an unused groupoid factor supplies no such evidence. These constructions should be described as optional structural interpretations, not a proved explanation of privacy or profit. No nontrivial higher coherence analysis of a production company was supplied.

A precise standard connection is available. In an ∞ -topos let $q : X \rightarrow Z$ be an effective epimorphism and put $R_n = X^{\times_{Z^{(n+1)}}$. Effective descent gives, for every target B ,

$$\mathrm{Map}(Z, B) \simeq \lim_{[n] \in \Delta} \mathrm{Map}(R_n, B). \quad (5)$$

Indeed, the augmented Čech nerve has colimit Z , and mapping out of that colimit yields (5) [L, sections 6.1-6.2]. For the finite constant objects used here, a channel with a **0-truncated** probability-vector target descends precisely when it is constant on quotient fibers. The response-aware criterion (2) verifies this obligation after marginalizing volatile responses. For a higher target, descent requires compatible homotopies and their coherences; the note did not construct or analyze those data.

Thus categorical descent expresses the finite criterion but does not strengthen it for free. The correction's technical content is the explicit operational test and counterexamples, not a new descent theorem. Remove the observation about the contractibility of ambient $\mathbb{R}^d$: the chosen quantized representation space is discrete, and the ambient homotopy type plays no role in these proofs. Numerical similarity alone remains insufficient to specify certified semantic equivalence.

6. Correct the parallel privacy grade

Execution concurrency does not determine privacy composition. Releases accessing overlapping protected users have the usual conservative sequential/adaptive bound $\sum_i \varepsilon_i$ when its conditional hypotheses hold, even if implemented at the same time.

For a **fixed public partition into disjoint protected-user sets**, with each mechanism accessing only its own part and using independent randomness, a neighboring change affects at most one component. All other probability factors are identical. The joint likelihood ratio is therefore bounded by

$$\exp\left(\max_i \varepsilon_i\right).$$

This is the usual disjoint-data parallel bound; it is not justified by disjoint request batches when a single protected user can occur in both. Neither arbitrary correlated randomness nor a privately chosen partition is covered by the stated proof. The correction distinguishes sequential addition from disjoint-user parallel maximum. The original additive parallel grade was only a conservative bound under appropriate composition hypotheses, not a sharp parallel rule. Cash accounting still requires its own assumptions and prefix-solvency test.

7. Correct the contribution and financing claims

Original label	Correct characterization
Theorem 3.1	Quotient factorization and uniqueness.
Proposition 3.2	Conditional induction; use criterion (2).
Theorem 4.1 and Proposition 4.2	Standard fiber criterion and MAP identity.
Theorem 5.1 and Proposition 5.2	Standard composition and output obstruction.
Theorem 6.1	Standard sheaf descent.
Theorem 7.1 and Corollary 7.2	Cash-prefix identity and rearrangements.
Theorem 8.1	Consistency example; payments assumed.

The financing calculation characterizes required revenue; it does not produce it. Retain the conditions

$$R \geq n(c - p) + F, \quad k_{\min} = \max \left(0, \max_j \sum_{t < j} d_t \right),$$

with the original definitions and realized event-ordering assumptions. Relabel the section 8 construction as a **worked example**, not a theorem deriving positive subsidy financing. Its asserted buyer payment remains an economic hypothesis. The synthetic subsidy of USD 9,000, net gain of USD 3,000, and opening liquidity of USD 15,000 remain arithmetically correct. No willingness to pay, retention implementation, or market viability has been established.

The defensible contribution is a finite compositional specification with explicit observation boundaries, a standard categorical descent interpretation, and conditional financial identities. A stronger higher categorical or empirical contribution requires further work.

8. Verification and authorship

The accompanying standard-library Python artifact checks the retry leak, criterion (2) on finite channels, the XOR metadata counterexample, a construction satisfying (4), a nonprivate charge appended to randomized response, valid postprocessed charges, disjoint-user versus shared-user privacy ratios, and the nonlocal-budget example. Exact rational arithmetic is used. It also reruns the original finite checks. These are finite examples, not proof-assistant certificates or a production privacy audit.

Authorship and assistance. Anonymous human author, responsible for the correction and its submission. OpenAI ChatGPT assisted with analysis, exposition, references, and computation. The correction was prompted by a user-supplied critical review. No identity, private contact information, or credentials are included. No independent referee acceptance is claimed.

References

- [O] Anonymous author. *Higher Categorical Semantics of an Embedding-Financed Proxy Company*. Higher Categorical Semantics, 25 October 2026. Editorially reviewed. <https://higher-categorical-semantics.pages.dev/research/higher-categorical-semantics-of-an-embedding-financed-proxy-company-9083e039>.
- [D] Cynthia Dwork and Aaron Roth. *The Algorithmic Foundations of Differential Privacy*. 2014. Proposition 2.1 and section 3.5. <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>.
- [L] Jacob Lurie. *Higher Topos Theory*. 2009, author’s revised text. Sections 6.1-6.2. <https://www.math.ias.edu/~lurie/papers/HTT.pdf>.