

Provenance-Aware Belief Updating in Human-AI Feedback Loops

Tianyi Huang
Ryquo

Eric Liu
App-In Club

Jay Chong
App-In Club

Emily Yin
App-In Club

Wanqi Zhao
App-In Club

tianyi@ryquo.com {eric.liu, jay.chong, emily, vicky}@appinclub.org

Abstract

In repeated human-AI interaction, a single observation can be quoted, summarized, and returned to a model as multiple reports, making repetition look like independent corroboration. We study whether language models can use supplied report lineage to update according to evidence provenance rather than report count. We introduce Latent Quotient Evidence Training (LQET), a relational objective that combines sparse Bayesian supervision with constraints that encourage stable predictions when reports add no new evidence and consistent updates to the same new evidence across relay histories. On a controlled binary benchmark, LQET reduces new-evidence update RMSE by 41-44% and log-odds drift after eight relay steps by 89-91% relative to task-only training across Qwen3-4B and Phi-4-mini, with three paired training seeds per backbone. The gains extend to a closed-loop simulation in which model confidence controls how many relayed reports re-enter the model’s context in later rounds. By grounding belief updates in supplied provenance, LQET offers a route toward dynamic human-AI systems in which influence follows new evidence, not the number of times the same evidence returns.

1 Introduction

A report is not always a new observation. In a coupled information system, one claim may be quoted by a model, copied into a summary, and returned through later context. A single speaker can sound like a chorus even when listeners know that repeated statements share a source [24]. In Bayesian social learning and decentralized estimation, inadvertent reuse of the same information (data incest) can bias beliefs and produce overconfidence [1; 15]. People also tend to neglect correlation among signals [7], while repeated human-AI interaction can amplify judgment biases through feedback [9]. Appropriate updating therefore depends not only on what reports say, but also on whether they introduce evidence or inherit it.

Prior work examines probabilistic updating from demonstrations [19], responses to evidence with varying informativeness and reliability [13], sequential belief trajectories [21], and sensitivity to repeated or paraphrased support [22; 17]. We ask a complementary structured question: given explicit multi-hop report provenance, can a model preserve its prediction when a report merely relays existing evidence, yet apply the same reliability-conditioned update when genuinely new evidence arrives after different relay histories? Sparse pointwise supervision on individual transcript views does not directly enforce this cross-transcript relationship: it assigns each transcript its own target, whereas the desired behavior concerns equivalence classes of transcripts and score changes between them.

We formalize this supplied-lineage setting with matched families of synthetic evidence reports. Prompts expose the derivation graph through parent-report links and natural-language descriptions of whether a report introduces an independent observation or relays an earlier one; evidence-root

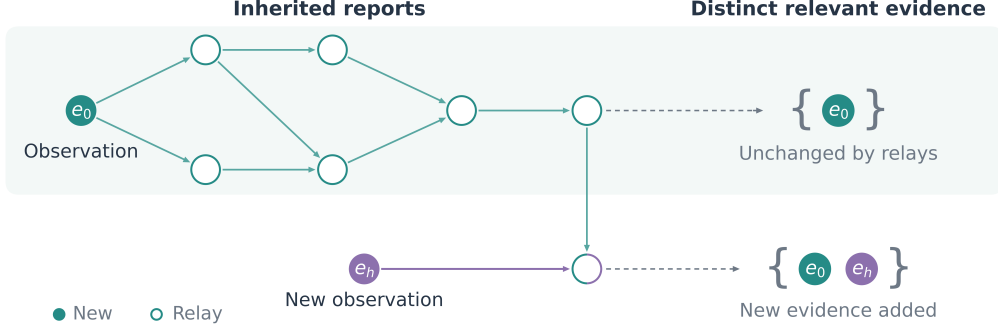


Figure 1: **Reports can multiply without adding evidence.** Relays preserve the evidence set $\{e_0\}$, while a new observation expands it to $\{e_0, e_h\}$. Solid arrows show report derivation; dashed arrows identify the distinct evidence in each history. With the prior and calibration records fixed, equivalent evidence sets have identical Bayesian targets.

identifiers used to construct the matched families are not rendered. Two transcripts are provenance-equivalent when they contain the same set of distinct, query-relevant evidence roots. Under the source-reliability model, exact Bayesian inference assigns the same log-odds target to every transcript in an equivalence class. The model should therefore satisfy two relations: adding only reports derived from existing evidence should leave its score unchanged, whereas adding a new evidence root should produce the same score change regardless of the preceding relay history.

Latent Quotient Evidence Training (LQET) directly supervises these relations. It combines sparse soft-target supervision with four invariance constraints, three new-evidence update-consistency constraints, and a presentation-order constraint. The learning problem is therefore to translate visible lineage into consistent belief updates across transcripts, rather than to discover unobserved report dependencies.

Our contributions are:

- We formalize provenance equivalence over sets of distinct evidence roots and derive exact Bayesian log-odds targets under uncertain source reliability.
- We introduce a relational objective that extends repetition invariance with new-evidence update consistency, composition, and presentation-order consistency while retaining sparse pointwise supervision.
- Across two language-model backbones, paired training seeds, and controlled distribution shifts, LQET consistently improves new-evidence update accuracy and stability across long relay histories.
- A confidence-dependent feedback simulation connects these static relations to the dynamic setting in which model confidence influences how many inherited reports return as later input.

2 Evidence provenance and Bayesian targets

Latent state and source reliability. Let $Y \in \{-1, +1\}$ be the binary answer with prior $\pi = P(Y = +1) \in (0, 1)$. Before observing calibration data, each source reliability R_s has an independent $\text{Beta}(1, 1)$ prior. Conditional on R_s , the source’s calibration outcomes are independent and identically distributed Bernoulli trials; a visible record of c_s correct and w_s incorrect measurements therefore gives

$$R_s \mid c_s, w_s \sim \text{Beta}(\alpha_s, \beta_s), \quad \alpha_s = c_s + 1, \quad \beta_s = w_s + 1. \quad (1)$$

An evidence root e denotes one distinct underlying observation, with reported outcome $x_e \in \{-1, +1\}$ and source $s(e)$. For a query-relevant root, $P(x_e = Y \mid Y, R_{s(e)}) = R_{s(e)}$. Conditional on Y and $\{R_s\}_s$, the outcomes of distinct roots are independent of one another and of the calibration outcomes. Multiple roots from the same source share R_s and thus become dependent after reliability is marginalized. By construction, conditional on the relevant-root observations and visible

calibration records, the report graph, rendering choices, and query-irrelevant roots carry no additional information about Y .

Reports and provenance equivalence. A report is a node in an acyclic derivation graph. It may link to parent reports, inherit their evidence roots, and introduce zero or more roots of its own. For transcript T , let

$$E(T) = \bigcup_{r \in T} \{e : e \text{ is a query-relevant root represented by } r\} \quad (2)$$

be the set of *distinct* relevant roots. Set union removes repeated descendants. Within a matched family, we define

$$T \sim T' \iff E(T) = E(T'). \quad (3)$$

This relation partitions each matched family into provenance-equivalence classes and induces the quotient used by LQET. To compute the Bayesian target, let $n_{s,a}(T) = |\{e \in E(T) : s(e) = s, x_e = a\}|$ for $a \in \{-1, +1\}$.

Exact Bayesian target. For $y \in \{-1, +1\}$, let

$$(C_s(y, T), W_s(y, T)) = \begin{cases} (n_{s,+}(T), n_{s,-}(T)) & y = +1, \\ (n_{s,-}(T), n_{s,+}(T)) & y = -1. \end{cases} \quad (4)$$

Let $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ and $\sigma(u) = (1 + e^{-u})^{-1}$. Write $\mathcal{O}(T) = ((s(e), x_e))_{e \in E(T)}$ for the indexed collection of relevant-root observations. After conditioning on the calibration records, let $\ell_y(T)$ be the log prior plus log marginal likelihood for $Y = y$; it equals the posterior log-probability up to an additive normalizing constant that is independent of y . Integrating each R_s yields

$$\ell_y(T) = \log P(Y = y) + \sum_s [\log B(\alpha_s + C_s(y, T), \beta_s + W_s(y, T)) - \log B(\alpha_s, \beta_s)], \quad (5)$$

$$t(T) = \ell_{+1}(T) - \ell_{-1}(T), \quad p^*(T) = P(Y = +1 \mid \mathcal{O}(T), \{c_s, w_s\}_s) = \sigma(t(T)). \quad (6)$$

Thus a relay that adds no new evidence leaves $t(T)$ unchanged, whereas a new root changes it according to its source, reported outcome, and the other roots from that source. The Bayesian target is constant within each provenance-equivalence class and, holding the prior and calibration records fixed, depends on a root set only through its source-by-outcome counts under the exchangeable source model.

Matched transcript families. Each static benchmark family produces 14 matched transcript views. The core 2×2 block crosses a relay history with the addition of new evidence: h_{00} contains the base root, h_{10} adds only relays, h_{01} adds a new root, and h_{11} adds that root after the relay history. Hence

$$t(h_{10}) = t(h_{00}), \quad t(h_{11}) = t(h_{01}), \quad t(h_{11}) - t(h_{10}) = t(h_{01}) - t(h_{00}). \quad (7)$$

A composition block sharing $c_{00} = h_{00}$ separately adds roots e_a and e_b in c_{10} and c_{01} , adds both in c_{11} , and reverses their presentation in c_{11}^{rev} . The base root and both added roots are drawn from three mutually distinct sources. Matched controls hold report author and displayed outcome fixed while varying the supplied provenance: a report either relays an existing observation or introduces a new one. Prompts include priors, calibration records, source names, report text, and explicit parent-report links; internal evidence-root and family identifiers are never rendered. A deterministic prompt-only calculator recovers every static-view Bayesian target, confirming that each target is fully specified by the visible prompt (Appendix E).

3 Latent Quotient Evidence Training

Forced-choice score. We evaluate a next-token forced-choice diagnostic rather than parse a generated numeric response. For prompt T , the score is the YES-NO log-odds

$$z_\theta(T) = L_\theta(\text{YES} \mid T) - L_\theta(\text{NO} \mid T), \quad \hat{p}_\theta(T) = \sigma(z_\theta(T)). \quad (8)$$

Here $L_\theta(a \mid T)$ is the pre-softmax next-token logit for token a . Thus $\hat{p}_\theta(T) = P_\theta(\text{YES} \mid T, \text{next token} \in \{\text{YES}, \text{NO}\})$: it is a forced-choice YES score, not the full-vocabulary probability of an unconstrained response. Each tokenizer maps YES and NO to distinct single tokens at the assistant boundary.

Sparse pointwise supervision. Seven views receive pointwise soft targets. Writing

$$\mathcal{A} = \{h_{00}, h_{01}, c_{10}, c_{01}, \text{same-source}, \text{counter}, \text{matched-new}\},$$

let N be the number of matched families in a training batch and define $p_{i,v}^* = p^*(T_{i,v})$ and $\hat{p}_{i,v} = \hat{p}_\theta(T_{i,v})$. Their average Bernoulli cross-entropy is

$$\mathcal{L}_{\text{task}} = -\frac{1}{N|\mathcal{A}|} \sum_{i=1}^N \sum_{v \in \mathcal{A}} [p_{i,v}^* \log \hat{p}_{i,v} + (1 - p_{i,v}^*) \log(1 - \hat{p}_{i,v})]. \quad (9)$$

Views such as h_{11} and c_{11} are supervised only through relational losses and receive no pointwise loss.

Relational constraints. For family i , write $z_i(v) = z_\theta(T_{i,v})$ and $t_i(v) = t_\theta(T_{i,v})$. We use the Huber penalty with $\delta = 0.25$,

$$\rho_\delta(u) = \begin{cases} \frac{1}{2}u^2, & |u| \leq \delta, \\ \delta(|u| - \frac{1}{2}\delta), & |u| > \delta. \end{cases} \quad (10)$$

The four provenance-invariance residuals are

$$\begin{aligned} r_{i,1}^{\text{inv}} &= z_i(h_{10}) - z_i(h_{00}), & r_{i,2}^{\text{inv}} &= z_i(h_{11}) - z_i(h_{01}), \\ r_{i,3}^{\text{inv}} &= z_i(\text{matched-copy}) - z_i(h_{00}), & r_{i,4}^{\text{inv}} &= z_i(\text{irrelevant}) - z_i(h_{00}), \end{aligned}$$

and their mean loss is

$$\mathcal{L}_{\text{inv}} = \frac{1}{4N} \sum_{i=1}^N \sum_{k=1}^4 \rho_\delta(r_{i,k}^{\text{inv}}). \quad (11)$$

The three new-evidence update residuals are

$$\begin{aligned} r_{i,1}^{\text{upd}} &= [z_i(h_{11}) - z_i(h_{10})] - [t_i(h_{01}) - t_i(h_{00})], \\ r_{i,2}^{\text{upd}} &= [z_i(c_{11}) - z_i(c_{10})] - [t_i(c_{01}) - t_i(c_{00})], \\ r_{i,3}^{\text{upd}} &= [z_i(c_{11}) - z_i(c_{01})] - [t_i(c_{10}) - t_i(c_{00})], \end{aligned}$$

with

$$\mathcal{L}_{\text{update}} = \frac{1}{3N} \sum_{i=1}^N \sum_{k=1}^3 \rho_\delta(r_{i,k}^{\text{upd}}). \quad (12)$$

The target differences in the latter two residuals use the benchmark’s three-way source distinctness: the base root and the two added roots come from mutually distinct sources. Updates from multiple roots sharing a source need not be additive. Finally, $\mathcal{L}_{\text{order}} = N^{-1} \sum_i \rho_\delta(z_i(c_{11}) - z_i(c_{11}^{\text{rev}}))$. The implemented objective is

$$\boxed{\mathcal{L}_{\text{LQET}} = \mathcal{L}_{\text{task}} + 4\mathcal{L}_{\text{inv}} + 2\mathcal{L}_{\text{update}} + \mathcal{L}_{\text{order}}.} \quad (13)$$

Each loss component is averaged separately before weighting. LQET uses 13 training views, whereas task-only uses the seven pointwise-supervised views; both methods use identical families and optimizer-update counts.

4 Experimental design

Backbones and training. We evaluate Qwen3-4B-Instruct-2507 [26; 20] and Phi-4-mini-instruct [16]. For each backbone and method, training seeds 0, 1, and 2 use the same fixed families: 4,000 for training and 400 for validation. We train for 500 updates (two epochs) with QLoRA [11; 4]: NF4 with double quantization, BF16 computation, LoRA rank 32, scaling factor $\alpha = 64$, dropout 0.05, and adapters on all linear modules. AdamW uses learning rate 2×10^{-4} , zero weight decay, a 3% warmup followed by cosine decay, gradient accumulation 16, and clipping at 1.0. Sequence length is capped at 768, and no evaluated prompt is truncated. The 1:4:2:1 loss-weight ratio, Huber threshold, and optimization settings were fixed before the reported Qwen runs and transferred unchanged to Phi; no validation-set hyperparameter search was conducted. Validation results served as run diagnostics and selected neither the final configuration nor checkpoints. We evaluate the final checkpoint.

Evaluation sets. Each run is evaluated on 400 independent and identically distributed (IID) test families from the training distribution, 800 controlled-shift out-of-distribution (OOD) families (160 for each of five factor groups: held-out domains and source names, longer relay histories, fan-in and diamond provenance graphs, held-out templates, and held-out prior and calibration-count supports), and 200 recursive-relay families at depths 1, 2, 4, 8. All 14 views are scored for the IID and OOD families; the recursive-relay families are evaluated on nested relay histories and their subsequent new-evidence probes. For the static IID and OOD views, task-only and LQET are scored on identical rendered prompts. Recursive-relay prompts are model-specific because each appended relay quotes that model’s preceding forced-choice score, although the underlying family, lineage schedule, and rendering rule are aligned. All partitions are mutually disjoint, and we check for overlap in sampled benchmark instances, evidence roots, family identifiers, and rendered prompts.

Primary endpoints. For the primary Qwen evaluation and Phi replication, new-evidence update RMSE and supervised-view KL pool the 800 families across the five equally sized OOD shifts within each seed; shift-specific estimates are secondary. New-evidence update RMSE compares predicted log-odds changes with the corresponding exact Bayesian changes on the three update pairs in Equation 12. Depth-8 relay drift averages the absolute score change from the base transcript after eight recursively generated relays over the 200 relay families. Supervised-view KL compares the exact Bayesian target with the forced-choice Bernoulli score and averages $D_{\text{KL}}(\text{Bern}(p^*) \parallel \text{Bern}(\hat{p}))$ over the seven pointwise-supervised views. Each backbone-level point estimate is the arithmetic mean of its three seed-level estimates; for RMSE, the square root is taken within each seed before averaging.

Statistical inference and decision criterion. We use 10,000 paired crossed seed-by-family bootstrap replicates. For each endpoint, a replicate independently resamples the three paired training seeds and one shared evaluation-family roster; the same family resample is reused across every selected seed occurrence and both methods, with all views from a family kept together. Pooled OOD resampling is stratified over the five fixed 160-family shifts, and nonlinear endpoints are recomputed within seed before averaging across resampled seeds. The 2.5th and 97.5th percentiles define unadjusted 95% intervals (Appendix C). The prespecified conjunctive criterion requires (i) negative LQET-minus-task-only effects for update RMSE and depth-8 relay drift in every seed, with pooled 95% upper bounds below zero, and (ii) an LQET-minus-task-only supervised-view KL effect of at most 0.02 in every seed, with a pooled upper bound of at most 0.02. This prespecified +0.02-nat non-inferiority margin for mean pointwise prediction quality was fixed before training and was not selected from the reported test results. We analyze each backbone separately and do not form a cross-backbone interval.

5 Results

5.1 Controlled results across backbones

Table 1 shows the same pattern on both backbones. LQET lowers new-evidence update RMSE by 41.25% on Qwen and 44.45% on Phi. Depth-8 relay drift falls by 91.36% and 89.38%, respectively. Both improvements hold in every training seed, and the three-seed mean update RMSE is lower on every OOD shift for both backbones. Pooled supervised-view KL changes by +0.00228 on Qwen and +0.00120 on Phi, with both upper bootstrap bounds below the prespecified +0.02-nat non-inferiority margin. Both backbones therefore satisfy the full conjunctive criterion. The IID test set shows the same directional pattern, with lower update RMSE and supervised-view KL on both backbones (Appendix D, Table 7).

Figure 2 breaks down the aggregate results by relay depth and distribution shift. Along nested relay histories, task-only drift accumulates sharply through depth four and remains near one log-odds unit at depth eight, whereas the three-seed mean for LQET remains below 0.14 at every depth (panels a,b). Update RMSE is lower on every controlled shift (panel c), with the largest proportional gains on unseen depth and topology and smaller gains on held-out prior and calibration supports. Supervised-view KL is lower or essentially unchanged on the other four shifts but increases on the held-out prior-and-calibration shift, the only shift that changes the numerical supports for both the class prior and source-calibration records (panel d). Each recursive relay quotes the preceding forced-choice

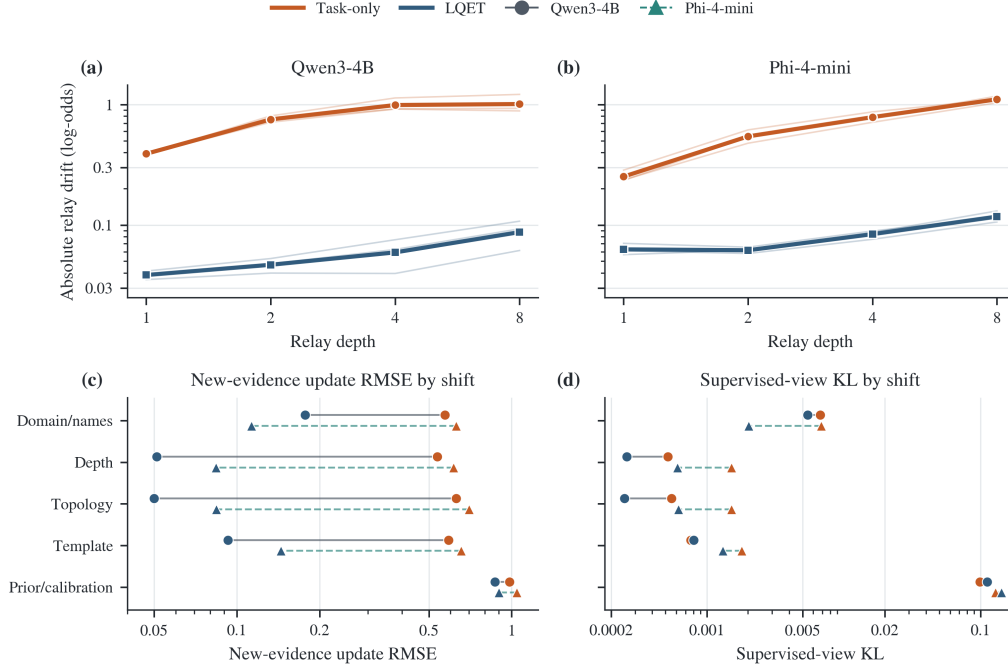


Figure 2: **Controlled generalization across relay depth and distribution shift.** (a,b) Absolute relay-induced drift across nested depths; thick and thin curves show three-seed means and individual seeds on the same 200 families. (c) Update RMSE and (d) supervised-view KL across controlled shifts; color marks the method, shape the backbone, and connectors the paired contrast. Panels (a,b) use logarithmic y-axes, and panels (c,d) use logarithmic x-axes. Lower is better.

YES score rounded to three decimals, so the depth test repeatedly introduces model-generated surface content rather than duplicating a fixed string.

Table 1: **Primary controlled results.** Effects are LQET minus task-only and are computed from unrounded estimates; lower is better. Point estimates average three paired training seeds. Intervals are 95% paired crossed seed-by-family bootstrap intervals. † marks the supervised-view KL endpoint with a prespecified +0.02-nat non-inferiority margin.

Backbone	Endpoint	Task-only	LQET	Effect [95% interval]
Qwen3-4B	Update RMSE	0.6816	0.4005	-0.2812 [-0.3070, -0.2565]
	Depth-8 drift	1.0122	0.0875	-0.9247 [-1.1175, -0.7703]
	Supervised-view KL †	0.0215	0.0237	+0.0023 [+0.00116, +0.00356]
Phi-4-mini	Update RMSE	0.7461	0.4144	-0.3317 [-0.3545, -0.3091]
	Depth-8 drift	1.1079	0.1177	-0.9902 [-1.0823, -0.8988]
	Supervised-view KL †	0.0280	0.0292	+0.0012 [-0.00002, +0.00251]

5.2 Selective invariance, not uniform update suppression

A model could lower relay drift by suppressing all updates. Matched copy-versus-new-evidence controls test this alternative directly (Figure 3a). LQET reduces base-to-relay and matched-copy drift by approximately 15-40 fold, while new-root and matched-new RMSE remain close to task-only. In the tightly matched comparison, which holds the report author and displayed outcome fixed while varying supplied provenance, copy-versus-new RMSE drops from 1.0145 to 0.4383 on Qwen and from 0.9818 to 0.4630 on Phi. The corresponding zero-intercept copy-versus-new slopes rise from 0.027 to 0.782 and from 0.067 to 0.769. Within this supplied-lineage setting, the pattern is selective: the model becomes less sensitive to declared relays while retaining its response to declared new observations.

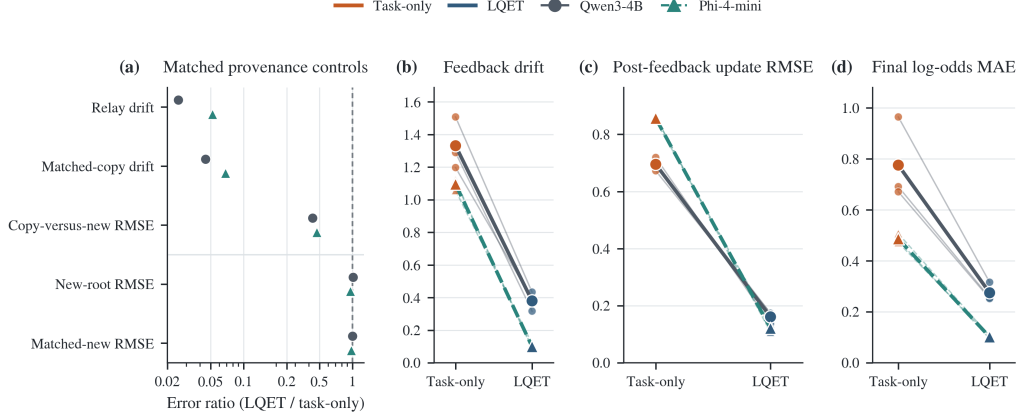


Figure 3: **Selective invariance and closed-loop feedback.** (a) LQET/task-only error ratio on matched provenance controls, shown on a logarithmic x-axis; unity denotes no change, and marker and color identify the backbone. Inherited-report and copy-versus-new errors fall while new-evidence update errors stay near parity. (b-d) Thin segments pair matched adapters by seed; thick segments and larger markers are three-seed means. Endpoint color marks the method, while shape and line style mark the backbone. All panels are descriptive; lower is better.

5.3 Closed-loop feedback dynamics

We next place each trained model in a deterministic human-AI feedback stress test on a fixed 64-family subset of the recursive-relay split (Appendix F). Crossing each family with initial trust $\tau_0 \in \{0.2, 0.8\}$ yields 128 trajectories per trained model. In round r , score-derived confidence c_r , trust τ_r , and the number of simulated human relays m_r obey

$$c_r = |2\sigma(z_r) - 1|, \quad \tau_{r+1} = 0.6\tau_r + 0.4c_r, \quad m_r = 1 + \min\{2, \lfloor 3\tau_{r+1}c_r \rfloor\}.$$

Each of three rounds appends one AI-derived report and $m_r \in \{1, 2, 3\}$ simulated human relays that add no new evidence. A fresh root is then added to measure the subsequent update.

All three metrics improve in every backbone-seed pair (Figure 3b-d). Mean reductions range from 64.6-76.9% on Qwen and 79.1-90.9% on Phi. This result connects the static provenance constraints to a dynamic failure mode: when confidence changes how many derived reports return, lower sensitivity to inherited reports also reduces feedback amplification.

5.4 What each relation contributes

A Qwen seed-0 component study evaluates each variant on fresh, disjoint sets of 800 controlled-shift families and 200 recursive-relay families. Full LQET and the no-order variant use the same 13 views and forward passes. Compared with no-order, full LQET lowers the order-sensitivity gap from 0.0426 to 0.0359, update RMSE from 0.4105 to 0.4058, and supervised-view KL from 0.0239 to 0.0232; both variants retain low depth-8 relay drift (0.0959 and 0.0753). The order term therefore contributes most clearly to presentation-order robustness. Comparing no-order with Pair-invariance suggests that update-consistency supervision together with the added composition views accounts for most of the compositional gain.

6 Related work

Belief updating and predictive consistency. Bayesian teaching can improve probabilistic reasoning in language models [19]. Evaluations study responses to evidence reliability [13], sequential belief trajectories [21], and martingale conditions for in-context learning [8]; Martingale Amortization aligns short- and longer-horizon predictions through self-consistency [14]. We instead ask whether supplied dependence changes the evidential force of otherwise matched reports.

Repetition bias and dependent social learning. Schuster et al. show that repetition can reverse language-model source preferences and reduce the effect by training on aligned repeated and unre-

peated contexts [22]; this setup most closely overlaps our invariance component. Naphade finds that paraphrases can outweigh genuinely distinct support and documents presentation-order effects [17]. Related work studies the chorus effect [24], persuasion bias [3], and correlation neglect [7]. Graph-based methods correct data incest when topology and the probabilistic model are available [1; 15], while Jain and Krishnamurthy illustrate data incest among interacting language-model agents [12]. Within our controlled setting, LQET extends repetition invariance to explicit multi-hop lineage, exact reliability-conditioned targets, new-root updates across relay histories, composition, and presentation order.

Feedback loops and evidence propagation. Language-model outputs can enter later contexts and create failures that static evaluation misses [18]; bidirectional alignment likewise emphasizes that model outputs shape later human inputs [23]. Multi-agent evaluation traces evidence propagation through later messages [25]. Our simulation isolates how score-dependent return of inherited reports interacts with provenance-sensitive updating.

Provenance, aggregation, and dependence discovery. Database provenance tracks information origins [2] and composes dependencies on input records [10]; natural-language claim-provenance graphs represent how claims originate and evolve [27]. GEAR is an adjacent evidence-aggregation method over graphs of retrieved sentences, not a report-lineage method [28]. Dong et al. infer copying and source dependence from output agreement [6] and update histories [5]. LQET assumes parent-report links are supplied and learns to map them to consistent belief updates, rather than retrieving evidence or discovering hidden dependence.

7 Scope and limitations

This study examines provenance-aware updating in a controlled binary setting with supplied parent links and a specified source-reliability model, enabling exact tests of evidence accounting. Evaluation across two 4B-class backbones and three paired seeds supports conclusions within this setting; transfer to larger models, naturalistic tasks, and incomplete or unreliable provenance remains for further study. The forced-choice YES-NO score assesses binary preference, while free-form probability estimation and downstream decision quality require separate evaluation.

Comparisons match training families and optimizer updates but differ in transcript-view exposure, so the reported gains characterize the complete training protocol. Under held-out prior and calibration supports, lower update RMSE accompanies higher supervised-view KL, identifying numerical generalization as a remaining challenge alongside improved relational consistency. The confidence-dependent simulation isolates report recirculation under a prescribed trust rule; studies with human participants are needed to assess how this mechanism operates in practical human-AI interaction.

8 Conclusion

LQET formulates evidence accounting as a relational learning problem, connecting predictions on matched transcripts through provenance invariance and new-evidence update consistency. Matched controls distinguish the learned stability from indiscriminate update suppression, while the confidence-dependent feedback simulation demonstrates reduced amplification when model predictions shape future inputs. Supervision thus extends from targets for individual transcripts to consistent updates across relay histories. Provenance gives this learning target a concrete structure: which changes in context warrant changes in belief.

References

- [1] T. Brehard and V. Krishnamurthy. Optimal data incest removal in bayesian decentralized estimation over a sensor network. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, volume 3, pages III–173–III–176, 2007. doi: 10.1109/ICASSP.2007.366500.
- [2] Peter Buneman, Sanjeev Khanna, and Wang Chiew Tan. Why and where: A characterization of data provenance. In *Proceedings of the 8th International Conference on Database Theory, ICDT '01*, page 316–330, Berlin, Heidelberg, 2001. Springer-Verlag. ISBN 3540414568.

- [3] Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118(3):909–968, 08 2003. ISSN 0033-5533. doi: 10.1162/00335530360698469. URL <https://doi.org/10.1162/00335530360698469>.
- [4] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 10088–10115. Curran Associates, Inc., 2023. doi: 10.52202/075280-0441. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/1feb87871436031bdc0f2beaa62a049b-Paper-Conference.pdf.
- [5] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. Truth discovery and copying detection in a dynamic world. *Proc. VLDB Endow.*, 2(1):562–573, August 2009. ISSN 2150-8097. doi: 10.14778/1687627.1687691. URL <https://doi.org/10.14778/1687627.1687691>.
- [6] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. Integrating conflicting data: the role of source dependence. *Proc. VLDB Endow.*, 2(1):550–561, August 2009. ISSN 2150-8097. doi: 10.14778/1687627.1687690. URL <https://doi.org/10.14778/1687627.1687690>.
- [7] Benjamin Enke and Florian Zimmermann. Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1):313–332, 01 2019. ISSN 0034-6527. doi: 10.1093/restud/rdx081. URL <https://doi.org/10.1093/restud/rdx081>.
- [8] Fabian Falck, Ziyu Wang, and Christopher C. Holmes. Is in-context learning in large language models bayesian? A martingale perspective. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 12784–12805. PMLR, 21–27 Jul 2024. URL <https://proceedings.mlr.press/v235/falck24a.html>.
- [9] Moshe Glickman and Tali Sharot. How human–ai feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 9(2):345–359, 2025. doi: 10.1038/s41562-024-02077-2. URL <https://doi.org/10.1038/s41562-024-02077-2>.
- [10] Todd J. Green, Grigoris Karvounarakis, and Val Tannen. Provenance semirings. In *Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS ’07, page 31–40, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936851. doi: 10.1145/1265530.1265535. URL <https://doi.org/10.1145/1265530.1265535>.
- [11] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=nZeVKeeFYf9>.
- [12] Adit Jain and Vikram Krishnamurthy. Interacting large language model agents. bayesian social learning based interpretable models. *IEEE Access*, 13:25465–25504, 2025. doi: 10.1109/ACCESS.2025.3538599.
- [13] Minsu Kim, Sangryul Kim, and James Thorne. From evidence to belief: A Bayesian epistemology approach to language models. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 10578–10611, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.naacl-long.531. URL <https://aclanthology.org/2025.naacl-long.531/>.
- [14] SongEun Kim, Seungyoo Lee, Edwin Fong, Hyungi Lee, and Juho Lee. From drift to coherence: Stabilizing beliefs in LLMs. In *Forty-third International Conference on Machine Learning*, 2026. URL <https://openreview.net/forum?id=0zkIq9YKml>.

- [15] Vikram Krishnamurthy and Maziyar Hamdi. Mis-information removal in social networks: Constrained estimation on dynamic directed acyclic graphs. *IEEE Journal of Selected Topics in Signal Processing*, 7(2):333–346, 2013. doi: 10.1109/JSTSP.2013.2245630.
- [16] Microsoft. Phi-4-mini-instruct. Hugging Face, 2025. URL <https://huggingface.co/microsoft/Phi-4-mini-instruct>.
- [17] Atharv Naphade. Rational synthesizers or heuristic followers? analyzing LLMs in RAG-based question-answering. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens, editors, *Findings of the Association for Computational Linguistics: ACL 2026*, pages 40293–40311, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.2003. URL <https://aclanthology.org/2026.findings-acl.2003/>.
- [18] Alexander Pan, Erik Jones, Meena Jagadeesan, and Jacob Steinhardt. Feedback loops with language models drive in-context reward hacking. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 39154–39200. PMLR, 21–27 Jul 2024. URL <https://proceedings.mlr.press/v235/pan24d.html>.
- [19] Linlu Qiu, Fei Sha, Kelsey Allen, Yoon Kim, Tal Linzen, and Sjoerd van Steenkiste. Bayesian teaching enables probabilistic reasoning in large language models. *Nature Communications*, 17:1238, 2026. doi: 10.1038/s41467-025-67998-6. URL <https://doi.org/10.1038/s41467-025-67998-6>.
- [20] Qwen Team. Qwen3-4B-Instruct-2507. Hugging Face, 2025. URL <https://huggingface.co/Qwen/Qwen3-4B-Instruct-2507>.
- [21] Ankur Samanta, Akshayaa Magesh, Tal Lancewicki, Ayush Jain, Youliang Yu, Paul Sajda, Kaveh Hassani, Aditya Modi, Daniel R. Jiang, and Yonathan Efroni. Bayesbench: Evaluating llm belief trajectories under multi-turn evidence accumulation, 2026. URL <https://arxiv.org/abs/2606.30850>.
- [22] Jakob Schuster, Vagrant Gautam, and Katja Markert. Whose facts win? LLM source preferences under knowledge conflicts. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens, editors, *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 29430–29459, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.1357. URL <https://aclanthology.org/2026.acl-long.1357/>.
- [23] Hua Shen, Tiffany Kneare, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Liu Yachuan, Savvas Petridis, Yi-Hao Peng, Li Qiwei, Chenglei Si, Yutong Xie, Jeffrey Bigham, Frank Bentley, Joyce Chai, Zachary Lipton, Qiaozhu Mei, Michael Terry, Diyi Yang, Meredith Morris, Paul Resnick, and David Jurgens. Position: Towards bidirectional human-ai alignment. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, Main Conference. Curran Associates, Inc., 2025. doi: 10.52202/085713-4186. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/b6ca96cdaad328fbee240e6fa8f609b0-Paper-Position_Paper_Track.pdf.
- [24] Kimberlee Weaver, Stephen M. Garcia, Norbert Schwarz, and Dale T. Miller. Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology*, 92(5):821–833, 2007. doi: 10.1037/0022-3514.92.5.821. URL <https://doi.org/10.1037/0022-3514.92.5.821>.
- [25] Chenfei Yan, Zeyang Yue, Feifei Zhao, Erliang Lin, Lu Jia, Haibo Tong, Mingyang Lyu, Chengyi Sun, and Yi Zeng. When truth is distributed: Misinformation derails collective fact recovery in llm-based multi-agent systems, 2026. URL <https://arxiv.org/abs/2608.03421>.

- [26] An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025. URL <https://arxiv.org/abs/2505.09388>.
- [27] Yi Zhang, Zachary Ives, and Dan Roth. “who said it, and why?” provenance for natural language claims. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4416–4426, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.406. URL <https://aclanthology.org/2020.acl-main.406/>.
- [28] Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. GEAR: Graph-based evidence aggregating and reasoning for fact verification. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 892–901, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1085. URL <https://aclanthology.org/P19-1085/>.

A Exact Bayesian target and benchmark construction

A.1 Beta–Bernoulli derivation

After conditioning on the public calibration record, fix $Y = y$ and a source s . Among the distinct relevant roots from s in $E(T)$, let $C_s(y, T)$ count outcomes that agree with y and $W_s(y, T)$ count outcomes that disagree. Integrating the source reliability in Equation 1 gives

$$P((x_e)_{e \in E(T): s(e)=s} \mid Y = y, c_s, w_s) = \frac{B(\alpha_s + C_s(y, T), \beta_s + W_s(y, T))}{B(\alpha_s, \beta_s)}. \quad (14)$$

Multiplying over sources and taking posterior log-odds yields Equation 6. The terms $\log B(\alpha_s, \beta_s)$ cancel between the two values of Y , giving

$$t(T) = \log \frac{\pi}{1 - \pi} + \sum_s [\log B(\alpha_s + n_{s,+}(T), \beta_s + n_{s,-}(T)) - \log B(\alpha_s + n_{s,-}(T), \beta_s + n_{s,+}(T))]. \quad (15)$$

Multiple distinct roots from one source are conditionally independent given (Y, R_s) but become dependent after R_s is marginalized. Descendant reports carrying the same root do not create repeated observations because $E(T)$ contains each root once.

The target is the exact Bayesian posterior under the benchmark’s binary source-reliability model. The `matched-new` and `counter` views are constructed controls rather than standard sampled views: the former is matched to the copied-report control in author and displayed outcome but declares a new measurement, whereas the latter displays the opposing outcome. Their targets are computed from the evidence declared in each prompt, allowing provenance and outcome polarity to vary separately.

A.2 Report provenance and validity

Every family is validated as a directed acyclic report graph. Each evidence root has exactly one introducing report; any later report carrying that root includes an explicit link to a parent report that also carries it. A pure relay introduces no root. A mixed report can inherit one or more roots and introduce a new root at the same node. Roots marked irrelevant to the query are excluded from $E(T)$. These rules make evidence multiplicity a property of provenance rather than the number of textual mentions.

A provenance equivalence class contains transcripts with the same set $E(T)$ of relevant evidence roots. Holding the prior and public calibration records fixed, equality of source-by-outcome counts is sufficient for the same Bayesian target under this benchmark, but it does not define equivalence-class membership: distinct roots from the same source with the same outcome remain distinct evidence.

B Matched transcript views

Table 2 lists the 14 matched views. Latent root labels such as e_0 , e_a , and e_b are construction notation and are never rendered in a prompt. LQET trains on 13 views; c_{00} is the same rendered transcript as h_{00} and contributes no distinct training term. The primary task-only and LQET runs are evaluated on all 14 views in the static IID and controlled-shift sets.

Table 2: **Matched transcript views.** *Supervised* denotes a pointwise soft target; *invariance* and *update* denote relational loss terms.

View	View construction	Bayesian-target relation	LQET role
h_{00}	Base observation carrying root e_0	base target	supervised; invariance/update reference
h_{10}	Base plus a relay DAG; no new root	$t(h_{10}) = t(h_{00})$	invariance; update reference
h_{01}	Base plus new root e_h	new-evidence target	supervised; invariance/update target
h_{11}	Relay DAG plus a mixed report introducing e_h	$t(h_{11}) = t(h_{01})$	invariance; update endpoint
c_{00}	Same rendered base as h_{00}	$t(c_{00}) = t(h_{00})$	static evaluation only
c_{10}	Base plus new root e_a	new-evidence target	supervised; update reference
c_{01}	Base plus new root e_b	new-evidence target	supervised; update reference
c_{11}	Base plus both e_a and e_b	updates add under source distinctness	update endpoint; order
c_{11}^{rev}	Same roots in reversed presentation order	$t(c_{11}^{\text{rev}}) = t(c_{11})$	order
matched-copy	Matched author/outcome; carries the base root through an explicit parent-report link	$t = t(h_{00})$	invariance
matched-new	Matched author/outcome; declares a new measurement and root	matched new-root target	supervised
same-source	New root from the base source	source-coupled target	supervised
irrelevant	Adds a root marked irrelevant to the query	$t = t(h_{00})$	invariance
counter	Adds a controlled root with the opposing outcome	controlled new-root target	supervised

The base root e_0 and the two composition roots e_a and e_b come from three distinct sources. The core-block root e_h is separate from these composition roots. Consequently,

$$t(c_{11}) - t(c_{00}) = [t(c_{10}) - t(c_{00})] + [t(c_{01}) - t(c_{00})]. \quad (16)$$

The benchmark does not assume that arbitrary additions from the same source are additive.

Table 3: **Training protocols.** Methods share training families and 500 optimizer updates.

Method	Training views	Objective
Task-only	7	$\mathcal{L}_{\text{task}}$
Pair-invariance	11	$\mathcal{L}_{\text{task}} + 4\mathcal{L}_{\text{inv}}$
No-order	13	$\mathcal{L}_{\text{task}} + 4\mathcal{L}_{\text{inv}} + 2\mathcal{L}_{\text{update}}$
Full LQET	13	Equation 13

Table 4: **Shared training configuration.**

Item	Value	Item	Value
Training families	4,000	Validation families	400
Dataset seed	17	Training seeds	0, 1, 2
Optimizer updates	500	Epochs	2
Families/microbatch	1	Gradient accumulation	16
Optimizer	AdamW	Learning rate	2×10^{-4}
Weight decay	0	Warmup	15 updates
Schedule	cosine	Gradient clip	1.0
Maximum length	768	Truncated prompts	0
Quantization	NF4 with double quantization	Compute dtype	BF16
LoRA rank	32	LoRA α / dropout	64 / 0.05
LoRA targets	all linear modules	Huber δ	0.25
Loss weights	task/invariance/update/order: 1:4:2:1	Checkpoint	final update

C Training and evaluation details

C.1 Training views and objective variants

C.2 Optimization configuration

The loss weights, Huber threshold, and optimization settings were fixed before the reported multi-seed Qwen runs and transferred unchanged to Phi; no validation-set hyperparameter search was conducted. Validation results served as run diagnostics only. The final checkpoint is evaluated without early stopping or validation-based checkpoint selection.

C.3 Generator supports and prompt format

Table 5 specifies the discrete benchmark supports. Each controlled shift changes only the named factor group; all remaining factors use the training support.

Table 5: **Benchmark-generator supports.** Calibration entries are numbers of correct and incorrect historical calibration trials, (c, w) .

Factor	Training, validation, and IID support	Controlled held-out support
Domain and names	Four domains: beacon status, valve position, package authenticity, and server security; 26 source names	Three domains: soil contamination, sample radioactivity, and component magnetization; 26 disjoint source names
Report graph	Chains of depth 1–3; stars with 1–8 relays	Depth shift: chains of depth 4–8. Topology shift: diamonds with 3 or 4 relays; fan-in graphs with branch depth 1 or 2 (3 or 5 relays total)
Rendering template	Direct and contrast renderings	Record-style and basis-style renderings
Prior π	$\{0.20, 0.35, 0.50, 0.65, 0.80\}$	$\{0.125, 0.275, 0.425, 0.575, 0.725, 0.875\}$
Calibration (c, w)	$c \in \{4, \dots, 9\}$ and $w \in \{1, 2, 3\}$	$\{(0, 0), (0, 3), (1, 4), (2, 0), (2, 5), (4, 4), (4, 10), (10, 0), (10, 4), (10, 10), (11, 0), (11, 5)\}$

Each family has four calibrated evidence sources; additional names can appear only as relay authors and contribute no independent observation. The prior and each calibrated source’s record are sampled uniformly from the applicable supports. For generatively sampled evidence roots, the generator draws $R_s \sim \text{Beta}(c_s + 1, w_s + 1)$, samples Y with $P(Y = +1) = \pi$, and draws an outcome that agrees with Y with probability R_s . Separate SHA-256-derived random-number streams govern evidence, calibration, domain, graph, and surface choices. A fixed SHA-256 rule assigns candidate family seeds to training, validation, and test partitions in an 80/10/10 ratio. Automated checks reject any cross-partition overlap in family identifiers, complete benchmark states, latent evidence configurations, or rendered prompts.

The benchmark intentionally supplies report lineage. In this representative prompt from the training distribution, parent citations and report text identify an independently acquired observation and its relay; the evaluated task is to use these declarations consistently across matched transcripts and report graphs.

Question: Is the valve open?

Prior probability of YES: 0.650

Calibration records:

- Ava previously made 8 correct and 2 incorrect measurements.
- Ben previously made 5 correct and 1 incorrect measurements.
- Dara previously made 6 correct and 3 incorrect measurements.
- Emil previously made 4 correct and 2 incorrect measurements.

Reports:

1. [R1] Ava: Cites: none. Ava independently ran a new position check without seeing any other report. The recorded outcome was OPEN.
2. [R2] Cleo: Cites: R1. Cleo read the cited predecessor report, ran no new measurement, and recorded the inherited outcome OPEN.

Instruction: Estimate the posterior probability that the answer is YES.

The displayed instruction is the verbatim instruction used in the experiments. We do not decode or parse a generated numeric answer. Instead, the reported quantity is the forced-choice YES score in Equation 8, conditional on the next token belonging to $\{\text{YES}, \text{NO}\}$. For each static IID or OOD view, the rendered prompt is identical across methods.

C.4 Static and recursive-relay evaluation sets

Each run is evaluated on a static 400-family IID set and five static controlled-shift sets of 160 families each:

- **domain/names:** held-out domains and source names;
- **depth:** relay depths 4–8;
- **topology:** diamond and fan-in report graphs;
- **template:** held-out rendering templates; and
- **prior/calibration:** held-out supports of priors and calibration counts.

All 14 matched views are scored on these static sets. Within each split, the ordered family roster is shared across methods, training seeds, and backbones.

A separate set of 200 recursive-relay families supports the depth and closed-loop feedback evaluations. For each family, one provenance-preserving chain is extended through depth 8; depths 1, 2, 4, and 8 are nested snapshots of that chain. The new-root probe at each snapshot is evaluated on a separate branch and is not included in later relay steps. The fixed ordered list of recursive-relay families, lineage schedule, and rendering rule is shared across methods, seeds, and backbones. Realized prompt text becomes model-specific after the first prediction because each appended relay quotes that model’s preceding forced-choice score. Training, validation, static evaluation, and recursive-relay partitions are mutually disjoint in benchmark instances, evidence roots, family identifiers, and rendered prompts. Every scored prompt fits within the length limit without truncation.

C.5 Endpoint definitions and bootstrap

For a fixed backbone and method, let seed $j \in \{0, 1, 2\}$ and suppress the backbone and method indices. Let $\mathcal{E}_{\text{upd}} = \{(h_{10}, h_{11}), (c_{10}, c_{11}), (c_{01}, c_{11})\}$ contain the three oriented new-evidence

update edges from Equation 12, written (v^-, v^+) . For family i and view v , write $z_{j,i}(v) = z_{\theta_j}(T_{i,v})$, $t_i(v) = t(T_{i,v})$, $p_i^*(v) = \sigma(t_i(v))$, and $\hat{p}_{j,i}(v) = \sigma(z_{j,i}(v))$. On the 800 pooled controlled-shift families, the new-evidence update RMSE is

$$\text{RMSE}_j^{\text{update}} = \sqrt{\frac{1}{800 \cdot 3} \sum_{i=1}^{800} \sum_{(v^-, v^+) \in \mathcal{E}_{\text{upd}}} ([z_{j,i}(v^+) - z_{j,i}(v^-)] - [t_i(v^+) - t_i(v^-)])^2}. \quad (17)$$

For recursive-relay family i , let $z_{j,i}^{\text{base}}$ be the base score and $z_{j,i}^{\text{relay},8}$ the score after eight provenance-preserving relays, before any new root is added. The seed-level depth-8 drift is

$$D_{8,j}^{\text{relay}} = \frac{1}{200} \sum_{i=1}^{200} |z_{j,i}^{\text{relay},8} - z_{j,i}^{\text{base}}|. \quad (18)$$

Let

$$\mathcal{A} = \{h_{00}, h_{01}, c_{10}, c_{01}, \text{same-source, counter, matched-new}\}$$

be the seven supervised views. Their seed-level supervised-view KL is

$$K_j^{\text{sup}} = \frac{1}{800|\mathcal{A}|} \sum_{i=1}^{800} \sum_{v \in \mathcal{A}} D_{\text{KL}}(\text{Bern}(p_i^*(v)) \parallel \text{Bern}(\hat{p}_{j,i}(v))). \quad (19)$$

For any seed-level endpoint q_j above, the reported backbone point estimate is $\bar{q} = \frac{1}{3} \sum_{j=0}^2 q_j$. In particular, each RMSE is square-rooted within its seed before the three seed-level values are averaged.

Because every trained model is evaluated on the same family rosters, training seeds and evaluation families form a crossed rather than nested design. Using random seed 20260829, each of 10,000 bootstrap replicates independently draws three paired seed labels with replacement and one common with-replacement family sample. For pooled controlled-shift endpoints, the family draw is stratified by resampling 160 families within each of the five fixed shifts; IID endpoints use one 400-family draw, and depth-8 drift uses one 200-family recursive-relay draw. The sampled family indices are reused across every selected seed occurrence and both methods, and all views or relay snapshots from a family remain grouped. We recompute each method-specific endpoint within seed, including taking the square root for RMSE, average over the three sampled seed occurrences, and form the LQET-minus-task-only contrast. We report the 2.5th and 97.5th percentiles without multiplicity adjustment.

D Primary results by seed and evaluation regime

D.1 Every training seed

Table 6: **Per-seed primary endpoints.** Effects are LQET minus task-only and are computed from unrounded estimates; lower is better.

Backbone	Endpoint	Seed	Task-only	LQET	Effect
Qwen3-4B	Update RMSE	0	0.6596	0.3901	−0.2695
		1	0.6895	0.4143	−0.2752
		2	0.6958	0.3971	−0.2987
	Depth-8 drift	0	1.2172	0.0931	−1.1241
		1	0.8875	0.1078	−0.7797
		2	0.9319	0.0616	−0.8703
	Supervised-view KL	0	0.0212	0.0226	+0.0014
		1	0.0227	0.0258	+0.0031
		2	0.0205	0.0229	+0.0024
Phi-4-mini	Update RMSE	0	0.7464	0.4074	−0.3391
		1	0.7439	0.4199	−0.3240
		2	0.7479	0.4160	−0.3319
	Depth-8 drift	0	1.0375	0.1156	−0.9219
		1	1.1775	0.1313	−1.0463
		2	1.1088	0.1063	−1.0025
	Supervised-view KL	0	0.0279	0.0281	+0.0001
		1	0.0277	0.0296	+0.0019
		2	0.0285	0.0301	+0.0015

D.2 IID test results

Table 7: **Results on the 400-family IID test set.** Point estimates average three paired training seeds. Effects are LQET minus task-only; intervals are 95% paired crossed seed-by-family bootstrap intervals. Lower is better.

Backbone	Endpoint	Task-only	LQET	Effect [95% interval]
Qwen3-4B	Update RMSE	0.5635	0.0515	−0.5119 [−0.5395, −0.4834]
	Supervised-view KL	0.00052	0.00023	−0.00029 [−0.00036, −0.00022]
Phi-4-mini	Update RMSE	0.6354	0.0865	−0.5489 [−0.5772, −0.5200]
	Supervised-view KL	0.00156	0.00061	−0.00094 [−0.00104, −0.00084]

D.3 Controlled-shift decomposition

Table 8: **Results by controlled shift.** Descriptive three-seed means are shown as task-only → LQET. Update RMSE is measured in log-odds units and KL in nats. Each shift has 160 families per seed; lower is better.

Backbone	Shift	Update RMSE	Supervised-view KL
Qwen3-4B	Domain/names	0.5717 → 0.1770	0.00673 → 0.00546
	Depth	0.5371 → 0.0511	0.00052 → 0.00026
	Topology	0.6287 → 0.0500	0.00055 → 0.00025
	Template	0.5903 → 0.0928	0.00076 → 0.00080
	Prior/calibration	0.9817 → 0.8694	0.09877 → 0.11196
Phi-4-mini	Domain/names	0.6278 → 0.1129	0.00685 → 0.00202
	Depth	0.6146 → 0.0839	0.00151 → 0.00061
	Topology	0.7009 → 0.0842	0.00151 → 0.00062
	Template	0.6556 → 0.1445	0.00180 → 0.00131
	Prior/calibration	1.0439 → 0.9005	0.12853 → 0.14164

The prior-and-calibration shift, which requires generalization to held-out numerical supports for both the class prior and source-calibration records, is the most difficult shift for both methods. LQET still lowers update RMSE from 0.9817 to 0.8694 on Qwen (11.4%) and from 1.0439 to 0.9005 on Phi (13.7%), although supervised-view KL rises by approximately 0.013 nats on each backbone. Across the other four shifts, supervised-view KL decreases in seven of eight backbone–shift comparisons and changes by only 4×10^{-5} in the remaining comparison.

D.4 Recursive-relay depth trajectory

Table 9: **Drift across recursive-relay depth.** Values are mean absolute pre-update log-odds drift over three seeds and correspond to Figure 2(a,b).

Backbone	Method	Depth 1	Depth 2	Depth 4	Depth 8
Qwen3-4B	Task-only	0.3911	0.7548	0.9932	1.0122
	LQET	0.0385	0.0469	0.0594	0.0875
Phi-4-mini	Task-only	0.2529	0.5454	0.7879	1.1079
	LQET	0.0629	0.0619	0.0842	0.1177

E Prompt sufficiency and selective-updating controls

E.1 Bayesian-target recovery from rendered prompts

A deterministic rule-based calculator receives only the rendered prompt. It parses the prior, calibration records, report text, relevance flags, and explicit parent-report links; it does not read Bayesian targets, hidden root identifiers, or family metadata. It successfully processes all $1,200 \times 14 = 16,800$ transcript-view instances in the static IID and controlled-shift sets. Its maximum absolute difference from the exact Bayesian log-odds computed directly from the benchmark state is below 5×10^{-15} , confirming that the Bayesian target is uniquely determined by the rendered prompt in this benchmark.

E.2 Matched copy-versus-new-root controls

Let $I = 800$ denote the pooled controlled-shift families for training seed j , and write $z_{j,i}(v)$ and $t_i(v)$ for the model and exact Bayesian log-odds of view v . For the six supervised new-root views

$$\mathcal{V}_{\text{new}} = \{h_{01}, c_{10}, c_{01}, \text{same-source}, \text{counter}, \text{matched-new}\},$$

define

$$r_{j,i}^{\text{new}}(v) = [z_{j,i}(v) - z_{j,i}(h_{00})] - [t_i(v) - t_i(h_{00})]. \quad (20)$$

The relay and matched-copy metrics are the mean absolute changes along their provenance-preserving edges,

$$D_{\text{relay},j} = \frac{1}{I} \sum_i |z_{j,i}(h_{10}) - z_{j,i}(h_{00})|, \quad (21)$$

$$D_{\text{copy},j} = \frac{1}{I} \sum_i |z_{j,i}(\text{matched-copy}) - z_{j,i}(h_{00})|. \quad (22)$$

The new-root RMSE averages all six supervised new-root views, whereas matched-new RMSE uses only the matched-new control:

$$\text{RMSE}_{\text{new},j} = \sqrt{\frac{1}{6I} \sum_i \sum_{v \in \mathcal{V}_{\text{new}}} r_{j,i}^{\text{new}}(v)^2}, \quad (23)$$

$$\text{RMSE}_{\text{matched-new},j} = \sqrt{\frac{1}{I} \sum_i r_{j,i}^{\text{new}}(\text{matched-new})^2}. \quad (24)$$

Finally, the copy-versus-new contrast holds report author and displayed outcome fixed while changing whether the report carries the base root through an explicit parent-report link or declares a new

measurement. With

$$\hat{\delta}_{j,i} = z_{j,i}(\text{matched-new}) - z_{j,i}(\text{matched-copy}), \quad (25)$$

$$\delta_i^* = t_i(\text{matched-new}) - t_i(\text{matched-copy}), \quad (26)$$

we report

$$\text{RMSE}_{\text{copy-new},j} = \sqrt{\frac{1}{I} \sum_i (\hat{\delta}_{j,i} - \delta_i^*)^2}, \quad (27)$$

$$\text{Slope}_j = \frac{\sum_i \delta_i^* \hat{\delta}_{j,i}}{\sum_i (\delta_i^*)^2}. \quad (28)$$

The slope is the zero-intercept regression coefficient of the predicted copy-versus-new change on the exact Bayesian change; its ideal value is one. Table 10 averages each seed-level metric over the three training seeds.

Table 10: **Provenance-selectivity metrics.** Values are descriptive means over three seeds. Lower is better except for the copy-versus-new slope, whose ideal value is one.

Backbone	Method	Relay drift	Copy drift	New-root RMSE	Matched-new RMSE	Copy-vs-new RMSE	Copy-vs-new slope
Qwen3-4B	Task-only	1.0463	0.8338	0.3980	0.4326	1.0145	0.027
	LQET	0.0264	0.0373	0.4043	0.4328	0.4383	0.782
Phi-4-mini	Task-only	1.0377	0.7813	0.4352	0.4682	0.9818	0.067
	LQET	0.0540	0.0533	0.4185	0.4582	0.4630	0.769

The `matched-copy` and `matched-new` views share the report author and displayed outcome; their supplied provenance differs because one relays the base observation and the other declares a new measurement. Because provenance status is intentionally explicit, this control tests whether the model uses supplied provenance while holding author and outcome fixed. Together with retained new-root update accuracy, the result argues against reliance on author/outcome alone and against uniform suppression of log-odds changes.

F Closed-loop feedback details

The closed-loop feedback analysis uses the first 64 families in the fixed ordered list of the same 200 recursive-relay families described in Section C. The subset rule and all simulator constants were fixed before evaluation. The two trust arms span low- and high-trust initial conditions, while three rounds with at most three relays per round define a bounded stress test; the same deterministic family prefix and settings preserve paired comparisons across methods, seeds, and backbones. Each family is evaluated at initial trust values 0.2 and 0.8, producing 128 trajectories per model and seed. If z_r is the current forced-choice log-odds, round r applies

$$p_r = \sigma(z_r), \quad c_r = |2p_r - 1|, \quad (29)$$

$$\tau_{r+1} = 0.6\tau_r + 0.4c_r, \quad m_r = 1 + \min\{2, \lfloor 3\tau_{r+1}c_r \rfloor\}. \quad (30)$$

Here c_r is confidence derived from the forced-choice score. The simulator appends one AI-derived report followed by $m_r \in \{1, 2, 3\}$ simulated human relays, then recomputes the model score from the rendered prompt. Each appended report includes an explicit link to its immediate parent report and introduces no new evidence root. After three rounds, the simulator appends an observation using the same new-root construction as h_{01} .

Let z_i^0 be the base log-odds, z_{ia}^{pre} the log-odds after three feedback rounds for family i and trust condition a , z_{ia}^{post} the log-odds after the new-root observation, and $t_{ia}^{\text{pre}}, t_{ia}^{\text{post}}$ the corresponding exact

Bayesian log-odds. With $M = 128$ trajectories, the endpoints are

$$D_{\text{feedback}} = \frac{1}{M} \sum_{i,a} |z_{ia}^{\text{pre}} - z_i^0|, \quad (31)$$

$$\text{RMSE}_{\text{postfb}} = \sqrt{\frac{1}{M} \sum_{i,a} [(z_{ia}^{\text{post}} - z_{ia}^{\text{pre}}) - (t_{ia}^{\text{post}} - t_{ia}^{\text{pre}})]^2}, \quad (32)$$

$$\text{MAE}_{\text{final}} = \frac{1}{M} \sum_{i,a} |z_{ia}^{\text{post}} - t_{ia}^{\text{post}}|. \quad (33)$$

Table 11: **Closed-loop feedback results by training seed.** Lower is better.

Backbone	Method	Seed	Feedback drift	Post-feedback update RMSE	Final log-odds MAE
Qwen3-4B	Task-only	0	1.5078	0.6936	0.9644
	LQET	0	0.4336	0.1732	0.3160
	Task-only	1	1.1973	0.6725	0.6909
	LQET	1	0.3887	0.1733	0.2560
	Task-only	2	1.2891	0.7191	0.6708
	LQET	2	0.3169	0.1362	0.2520
Phi-4-mini	Task-only	0	1.0566	0.8477	0.4717
	LQET	0	0.1172	0.1361	0.1041
	Task-only	1	1.1191	0.8601	0.5065
	LQET	1	0.0938	0.1197	0.1056
	Task-only	2	1.1113	0.8610	0.4810
	LQET	2	0.0869	0.1068	0.0947

LQET also reduces the mean cumulative number of simulated human relays from 5.4349 to 4.4557 on Qwen and from 5.0469 to 4.3438 on Phi. Because the number of relays depends on score-derived confidence, these trajectories capture both reduced sensitivity to provenance-preserving reports and the associated reduction in how many such reports return.

G Component analysis

The Qwen3-4B seed-0 component study uses fresh, disjoint evaluation sets: 800 static controlled-shift families (160 per shift) for update RMSE, supervised-view KL, and order-sensitivity gap, and 200 recursive-relay families for depth-8 drift. For the static controlled-shift families, define

$$G_{\text{order}} = \frac{1}{800} \sum_{i=1}^{800} |z_i(c_{11}) - z_i(c_{11}^{\text{rev}})|. \quad (34)$$

Full LQET and No-order both process 13 training views and use identical forward computation. No-order sets the order-loss weight to zero, so c_{11}^{rev} has no loss contribution. Pair-invariance processes 11 training views. Intervals use 10,000 paired family-bootstrap draws over the relevant fresh static or recursive-relay families.

Table 12: **Qwen3-4B seed-0 component results.** Lower is better.

Method	Update RMSE	Depth-8 drift	Supervised-view KL	Order gap
Pair-invariance	0.5860	0.0984	0.0212	0.1184
No-order	0.4105	0.0753	0.0239	0.0426
Full LQET	0.4058	0.0959	0.0232	0.0359

Table 13: **Paired component effects.** Effects are No-order minus full LQET and use unrounded estimates. Brackets are unadjusted 95% paired family-bootstrap intervals; positive effects favor full LQET for these lower-is-better endpoints.

Endpoint	Effect	95% CI
Update RMSE	+0.00475	[+0.00157, +0.00782]
Depth-8 drift	-0.02063	[-0.03125, -0.01031]
Supervised-view KL	+0.00076	[+0.00031, +0.00125]
Order gap	+0.00664	[+0.00289, +0.01023]

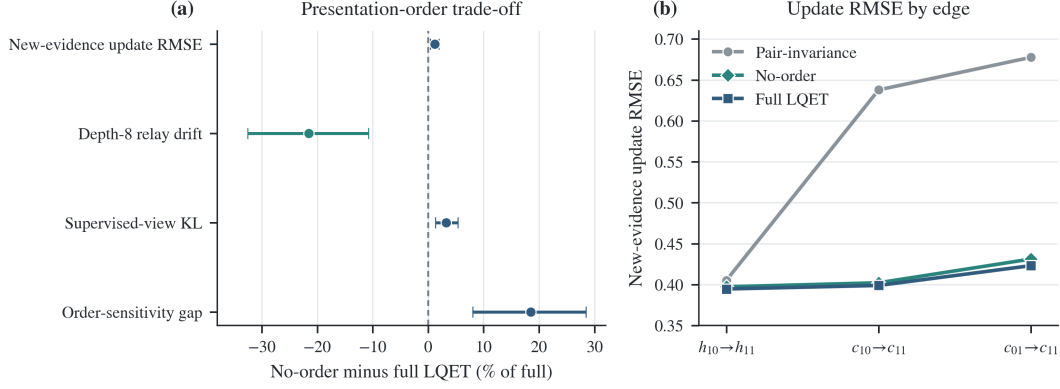


Figure 4: **Single-seed component trade-off.** (a) No-order-minus-full effects normalized by full LQET, with 95% paired family-bootstrap intervals; positive values favor full LQET for the displayed lower-is-better endpoints. (b) Edgewise update RMSE across component variants.

The order term yields its clearest improvement on the order-sensitivity gap and modest improvements in update RMSE and supervised-view KL, while No-order has lower depth-8 drift for this seed (Figure 4a). Relative to Pair-invariance, No-order reduces update RMSE by 0.1755 and the order-sensitivity gap by 0.0759; the two composition edges account for most of the update improvement (Figure 4b). Because adding update consistency also adds the composition views, this comparison estimates their combined contribution rather than isolating the loss term alone.