TASK-ALIGNED CONTEXT DISTILLATION: SET-WISE SELECTION WITH STRUCTURAL AND GEOMETRIC PRIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Long-context distillation aims to compress noisy inputs into minimal sufficient subsets while preserving downstream reasoning performance. Existing methods fall into two paradigms: *point-wise* methods score context units without modeling inter-chunk dependencies, while *heuristic set-wise* methods optimize diversity without grounding in solver feedback. Both are susceptible to semantic collapse, structure blindness, or solver mismatch under strict compression budgets.

We reframe context distillation as a **solver-aligned set-wise** optimization problem, and introduce **TaCo**, a solver-in-the-loop policy learning framework. Solving this formulation via naive reinforcement learning (RL) exposes a fundamental challenge: continuous embeddings provide no inductive bias, yielding discontinuous reward surface and high reward variance that renders credit assignment intractable. By identifying two independent sources of reward instability, TaCo overcomes this with two complementary priors: a **structural prior** preserves the internal integrity of each context unit, and a **geometric prior** promotes the diversity across selected units. Together, these priors substantially reduce the effective RL search space, enabling stable credit assignment under sparse solver feedback.

On academic long-context benchmarks, TaCo outperforms SOTA compressors on unstructured text and even exceeds the full-context baseline on structured text. On an industrial benchmark with extreme length, high noise and mixed structure, TaCo achieves 98% and 102% of full-context F1 at 2.18% and 9.71% token retention respectively. Beyond aggregate performance, mechanistic diagnostics confirm that TaCo's three core components—the structural prior, geometric prior, and solver-aligned policy—orthogonally mitigate all three failure modes, validating the representational design of the framework.

1 INTRODUCTION

Effective reasoning over long contexts demands selecting the right information. Despite rapid progress extending LLM context capacity beyond one million tokens, reasoning quality degrades as input volume grows: attention mechanisms are distracted by irrelevant content (Shi et al., 2023), and critical evidence buried in the middle of long sequences is systematically underutilized (Liu et al., 2024). This motivates long-context distillation: compressing a noisy input into a Minimal Sufficient Subset (MSS) that preserves downstream task performance under a strict token budget.

The majority of existing methods operate under a *point-wise* assumption: scoring context units independently (Jiang et al., 2023a; Lewis et al., 2020). Even recent query-aware compressors such as TPC (Cao et al., 2024) and Adaptive QuerySelect (Liskavets et al., 2025) largely inherit this mechanism. Under strict compression budgets, greedy point-wise selection inevitably induces two failure modes. *(1) Semantic Collapse*—over-sampling redundant chunks while omitting complementary signals required for multi-hop reasoning (Xu et al., 2024; Ram et al., 2023). *(2) Structure Blindness*—fragmenting logically indivisible units (e.g., a JSON object split across chunk boundaries) and disrupting structural dependencies prerequisite to correct reasoning (Jiang et al., 2023b; Wang et al., 2024b; Xie et al., 2022). These failures motivate reframing distillation as set-wise subset selection: choose $\tilde{\mathcal{D}}$ that jointly maximizes $\mathcal{U}(\mathcal{M}(\tilde{\mathcal{D}}, q), y)$ under token budget κ.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

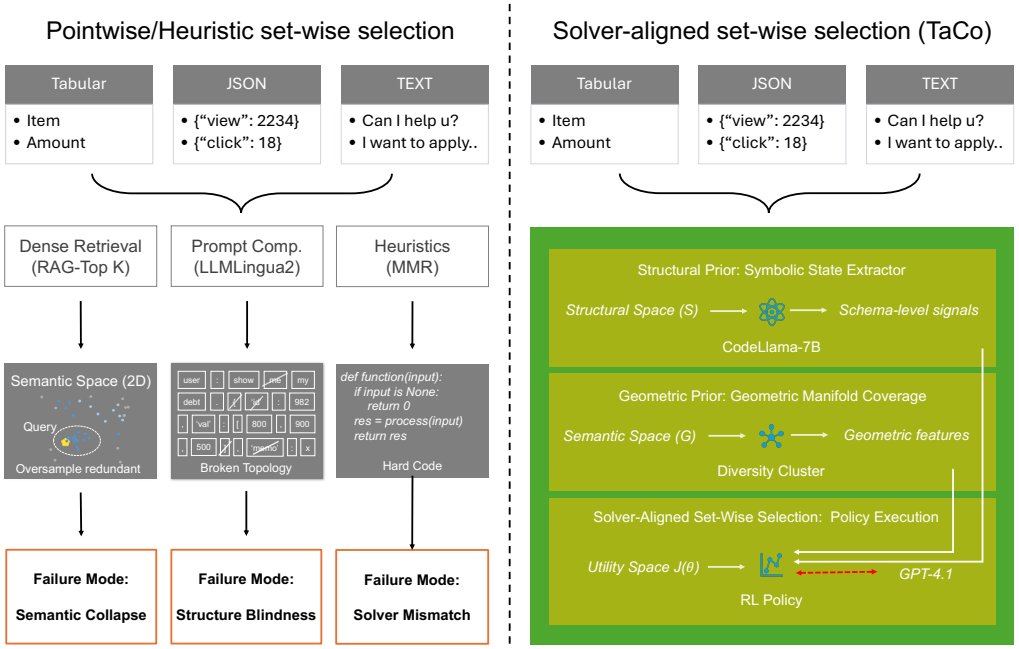


Figure 1: Left: three failure modes of existing context compression paradigms. Right: TaCo framework, where Symbolic and Manifold modules provide two inductive biases that address Structure Blindness and Semantic Collapse respectively, while the RL Policy mitigates Solver Mismatch.

Explicitly defining heuristic set-wise diversity rules (e.g., MMR (Carbonell & Goldstein, 1998)) introduces a third failure mode: *(3) Solver Mismatch*—subsets are selected without grounding in solver feedback, producing inputs that may be locally coherent yet misaligned with the target LLM's reasoning patterns, which can only be captured via direct feedback. This motivates solver-aligned set-wise optimization, where selection decisions are grounded in explicit solver feedback.

Employing a solver-in-the-loop RL policy is theoretically sound, but solving this delayed-reward combinatorial problem directly over raw contexts leads to severe optimization instability. Operating on high-dimensional continuous embeddings provides *no inductive bias* regarding document topology or information redundancy, inducing massive state-space variance and credit assignment difficulties under sparse rewards, making stable convergence intractable.

We propose **TaCo** (Fig. 1), a solver-aligned set-wise distillation framework that decouples state representation from policy learning. TaCo introduces two inductive biases: (1) Symbolic State Extraction operates on discrete, indivisible structural blocks extracted via schema-guided code generation, capturing schema-level signals to prevent fragmentation and mitigating Structure Blindness. (2) Geometric Manifold Coverage provides pre-computed distances to cluster centers as dense geometric features, translating globally redundant information into locally complementary signals and mitigating Semantic Collapse. Together, these modules substantially reduce the effective RL search space, enabling stable convergence to solver-aligned selection and mitigating Solver Mismatch.

Our contributions are: **(1) Problem Reformulation.** We expose the limitations of point-wise and heuristic set-wise paradigms, reformulate distillation as solver-aligned set-wise optimization, and provide a formal proof (Appendix A). **(2) Methodological Solution.** We show that naive solver-in-the-loop RL suffers from two independent sources of reward instability rooted in orthogonal information dimensions, and derive two corresponding inductive biases that make solver-aligned set-wise optimization tractable within a unified RL framework. **(3) Zero-Shot Generalization & Extreme Compression.** On academic long-context benchmarks, TaCo outperforms SOTA compressors on unstructured text and even exceeds the full-context baseline on structured text. On an industrial benchmark with extreme length, high noise and mixed structure, TaCo achieves 98% / 102% of full-context F1 at only 2.18% / 9.71% token retention. **(4) Mechanistic Verification.** Mechanistic diagnostics confirm that TaCo's three core components—the structural prior, geometric prior, and solver-aligned policy—orthogonally mitigate Structure Blindness, Semantic Collapse, and Solver Mismatch, respectively, validating the representational design of the framework.

Full code will be released upon acceptance; a reference implementation is provided for review.

2 Related Work

Point-Wise Context Compression and Retrieval. Architectural solutions such as sparse attention (Child et al., 2019; Beltagy et al., 2020; Zaheer et al., 2020) and KV-cache eviction (Xiao et al., 2024; Zhang et al., 2023) extend context capacity but do not address reasoning degradation from irrelevant content (Shi et al., 2023; Liu et al., 2024). Information-theoretic compressors (Selective Context (Li et al., 2023), LLMLingua series (Jiang et al., 2023a; 2024)) prune tokens via perplexity or self-information. Query-aware methods (TPC (Cao et al., 2024), Adaptive QuerySelect (Liskavets et al., 2025), RECOMP (Xu et al., 2024)) condition chunk relevance on the user query. LLMLingua-2 (Pan et al., 2024) distills LLM feedback into pseudo-labels for a token-level classifier. Context-Picker (Zhu et al., 2024) frames selection as identifying a minimal sufficient subset via reasoning-aware scoring. Despite their diversity, all these methods share a *point-wise* paradigm: scoring units independently under an additive utility assumption, which systematically neglects joint information structure and leads to semantic collapse and structure blindness (Sec. 3.1).

Heuristic Set-Wise Diversity Constraints. Subset selection techniques such as Coresets (Sener & Savarese, 2018; Mirzasoleiman et al., 2020) and DPPs (Kulesza & Taskar, 2012) have inspired diversity-aware retrieval. In RAG and summarization (Wang et al., 2024a), MMR enforces geometric diversity by penalizing semantic similarity among selected chunks. While these methods introduce a set-wise perspective, they suffer from geometric rigidity—spatial distance does not reliably correspond to logical complementarity—and remain entirely decoupled from downstream solver feedback, leaving solver mismatch unaddressed (Sec. 3.1).

Solver-in-the-Loop Optimization. RL with solver feedback (Williams, 1992) provides a theoretically rigorous alternative to static heuristics. HiLight (Li et al., 2026) demonstrates the viability of closed-loop solver feedback for context selection via token-level soft highlighting. However, by applying RL directly over raw embeddings, HiLight disperses credit across thousands of token-level decisions without structural or geometric inductive biases. ZipRL (Hu et al., 2026) attacks the sparse-reward credit-assignment problem from the reward side rather than the state side—densifying the reward signal while keeping the state representation fixed. TaCo instead restructures the state representation itself, making sparse, delayed rewards tractable without requiring hindsight relabeling. More broadly, decoupling the RL policy from raw embedding noise to make credit assignment tractable under sparse rewards remains an open challenge—one that TaCo directly targets (Sec. 3).

3 Method

3.1 Problem Formulation

Let $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ denote a sequence of atomic context units, q the user query, and $\mathcal{M}$ a frozen downstream solver. A binary selection vector $\mathbf{a} \in \{0,1\}^N$ induces a compressed subset $\tilde{\mathcal{D}}_{\mathbf{a}} = \{x_i \mid a_i = 1\}$ under token budget κ. The goal of solver-aligned set-wise context distillation is:

$$\max_{\mathbf{a}} \mathcal{U}\Big(\mathcal{M}(\tilde{\mathcal{D}}_{\mathbf{a}}, q), y\Big) \quad \text{s.t.} \quad \sum_i a_i \cdot \text{len}(x_i) \le \kappa, \tag{1}$$

where $\mathcal{U}$ is a task-level metric and y the ground-truth output. Appendix A proves an expressiveness gap in point-wise methods and a solver alignment gap in heuristic set-wise methods, motivating a solver-in-the-loop RL framework that directly optimizes set-level utility via solver feedback.

Two Independent Sources of RL Reward Instability. Real-world long-context documents mix two orthogonal dimensions of information, each exposing a naive RL policy to an independent source of reward instability: **structural information** encodes intra-unit discrete boundaries whose integrity is a precondition for task-relevant value; **semantic information** encodes inter-unit overlap and complementarity whose marginal utility depends on the full candidate pool. **(1) Structure Blindness → Discontinuous Reward Surface**: discrete boundaries are smoothed in continuous space—a complete record and its truncated variant satisfy $\|m - m'\|_2 \approx 0$ yet $|R(m) - R(m')| \gg 0$. **(2) Semantic Collapse → High Reward Variance**: inter-unit redundancy is invisible in point-wise

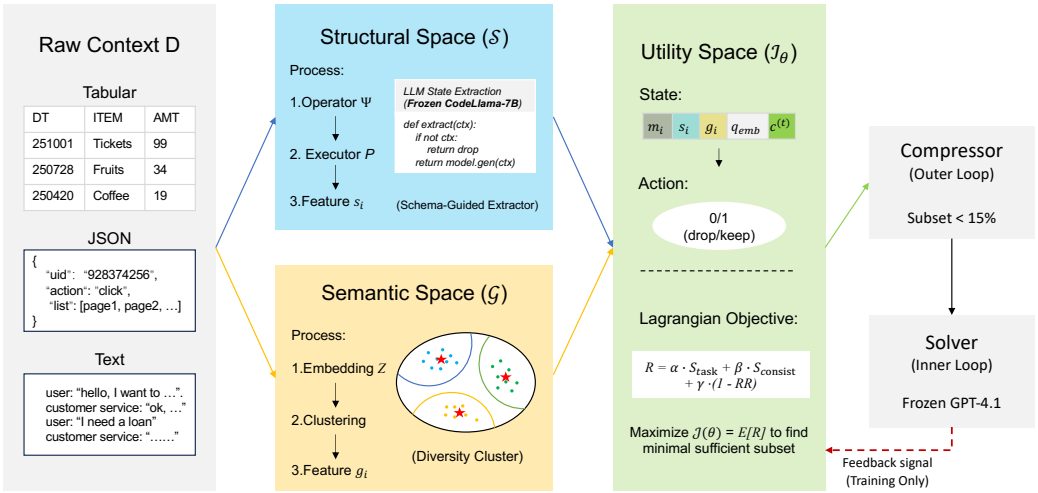


Figure 2: Architecture of TaCo. Inputs undergo symbolic state extraction and geometric manifold coverage before a policy network sequentially selects the final subset based on solver feedback.

encodings—two units occupying the same manifold region receive identical policy states despite near-zero marginal utility when co-selected. The two sources are rooted in orthogonal dimensions; no single inductive bias can suppress both (Appendix C.3, Figure 3).

TaCo's Solution. TaCo (Fig. 2) introduces two dedicated priors, one per dimension: a **structural prior** (Sec. 3.2) that makes discrete unit boundaries explicit before the policy acts, reducing reward surface discontinuities; and a **geometric prior** (Sec. 3.3) that encodes each unit's position on the semantic manifold as explicit policy state features, making inter-unit redundancy observable prior to policy execution and reducing conditional reward variance. **Policy execution** (Sec. 3.4) then grounds set-level selection in task utility via solver feedback. Table 7 in Appendix C.1 lists the configuration and hyperparameters for TaCo; end-to-end efficiency analysis is provided in Appendix E.1, Table 11.

3.2 Structural Prior : Symbolic State Extraction

Failure Mode Addressed. Under strict compression budgets, dropping a single schema key or closing bracket silently fragments a complete information unit, producing wildly inconsistent rewards for superficially similar actions.

Module Design. We decompose symbolic state extraction into two stages: a schema-conditioned code generation operator Ψ, realized via a frozen CodeLlama-7B-Instruct, and a code executor $P = \Psi(\mathcal{T})$ synthesized from the document schema $\mathcal{T}$ (once per schema type, amortized across all samples). For each unit x_i, the executor P is applied per sample to extract a key-value mapping $e_i = P(x_i)$; e_i is then serialized to a canonical string and embedded as $s_i = \text{Embed}(\text{serialize}(e_i)) \in \mathbb{R}^d$. (See an illustrative implementation of Ψ in Appendix C.2.)

Design Rationale. Symbolic state extraction acts as a deterministic discrete filter: The policy operates on discrete, indivisible structural blocks whose boundaries are enforced deterministically, suppressing the discontinuous reward surface. If x_i contains no extractable structured fields (e.g., plain prose without JSON, tabular structure or code blocks), P returns an empty mapping and $s_i = \mathbf{0}$, indicating the unit carries no symbolic structure.

3.3 Geometric Prior: Geometric Manifold Coverage

Failure Mode Addressed. Even after structural boundaries are enforced, point-wise encodings provide no signal about where each unit lies on the semantic manifold: two units in the same region receive identical policy states despite near-zero marginal utility when co-selected.

Module Design. Context units are embedded into a semantic manifold $\mathcal{Z}$ and clustered (precomputed, query-agnostically) with cluster count $K = \lfloor \rho \cdot \bar{\kappa}/\bar{l} \rfloor$, $\bar{\kappa}$ a length-proportional refer-

ence budget, $\bar{l}$ the mean unit length, and $\rho \geq 1$ an oversampling factor. For each cluster C_k, we cache its centroid μ_k. At inference time, for unit $x_i \in C_k$, the geometric feature vector $g_i = \mathbf{W}_g \left[\mathrm{sim}(x_i, q) \oplus \mathrm{sim}(x_i, \mu_k) \oplus \mathrm{sim}(x_i, \hat{x}_k)\right]$, where $\mathbf{W}_g \in \mathbb{R}^{d\times 3}$ is a learnable projection matrix, $\hat{x}_k$ the cluster medoid, and the three scalars encode query relevance, intra-cluster position, and local representativeness, respectively.

Design Rationale: Why Not MMR. MMR (Carbonell & Goldstein, 1998) penalizes redundancy via a flat pairwise similarity threshold, with no awareness of each unit's position on the global manifold, and cannot distinguish redundant units (same region of $\mathcal{Z}$) from complementary ones (distinct regions despite surface similarity). Geometric Manifold Coverage instead encodes manifold position as learnable policy state features, making inter-unit redundancy directly observable in $h_i^{(t)}$ and reducing $\mathrm{Var}[R \mid h_i^{(t)}]$ (Corollary B.1, Appendix B).

3.4 Solver-Aligned Set-Wise Selection: Policy Execution

Failure Mode Addressed. With both instability sources suppressed, the remaining challenge is *solver mismatch*: only a learned policy trained on solver feedback can ground selection in task-level performance for the solver it is trained against.

Module Design. A 12-layer Transformer Encoder (see capacity analysis in Appendix C.3, Table 9) observes per-unit state: $h_i^{(t)} = [m_i, s_i, g_i, q_{\mathrm{emb}}, c^{(t)}]$, where m_i is the raw embedding, s_i the symbolic state (Sec. 3.2), g_i the geometric feature (Sec. 3.3), q_{emb} the query embedding, and $c^{(t)} = \left(\sum_{j\in\mathcal{S}^{(t)}} m_j\right)/|\mathcal{S}^{(t)}|$ ($c^{(0)} = \mathbf{0}$) the selection context vector over already-selected units $\mathcal{S}^{(t)}$. At each step t, the policy scores all remaining units, selects one, and updates $c^{(t)}$; selection terminates when adding the next unit would exceed κ. Since inter-unit redundancy is handled upstream by g_i, the first-order set-level coverage signal provided by $c^{(t)}$ suffices for stable credit assignment under sparse solver feedback (Appendix C.3, Figure 3), while avoiding the $O(|\mathcal{S}^{(t)}|)$ overhead of attention-based aggregation. The policy outputs:

$$\pi_\theta(a_i = 1 \mid h_i^{(t)}) = \sigma\Big(\mathrm{MLP}\Big(\mathrm{Enc}_\theta(h_i^{(t)})\Big)\Big), \tag{2}$$

with $a_i \sim \mathrm{Bernoulli}(\pi_\theta)$ subject to κ. We instantiate $\mathcal{U}(\cdot) \triangleq R$ in Eq. (1), where the reward R decomposes into three complementary signals (see sensitivity analysis in Appendix C.3, Table 8):

$$R = \alpha \cdot \mathcal{S}_{\mathrm{task}} + \beta \cdot \mathcal{S}_{\mathrm{consist}} + \gamma \cdot (1 - \mathrm{RR}), \tag{3}$$

where $\mathcal{S}_{\mathrm{task}}$ (F1/ROUGE-L) measures task performance, $\mathcal{S}_{\mathrm{consist}}$ (cosine similarity between the sentence embeddings of solver output and ground truth using BGE-M3) measures output fidelity, and $\mathrm{RR} = \sum_i a_i \cdot \mathrm{len}(x_i) \;/\; \sum_i \mathrm{len}(x_i)$ the token retention rate ($\alpha = 0.25$, $\beta = 0.40$, $\gamma = 0.35$).

Optimization. We instantiate the solver-in-the-loop policy via REINFORCE (Williams, 1992; Sutton & Barto, 2018) with a moving average baseline, stabilized by entropy regularization ($\lambda_H = 0.01$). Selection order is randomized per episode to prevent order overfitting. Convergence diagnostics are in Appendix C.3, Figure 3.

4 Experiments

We evaluate TaCo through three research questions: zero-shot transfer (RQ1, Sec. 4.2), extreme industrial compression (RQ2, Sec. 4.3), and ablation with failure mode verification (RQ3, Sec. 4.4).

4.1 Experimental Setup

Datasets. We evaluate on three benchmark suites. **(1) LongBench** (Bai et al., 2024): benchmarks covering multi-hop reasoning (HotpotQA), summarization (QMSum), and scientific QA (Qasper), evaluating zero-shot transfer against representative compressors. **(2) Native long-context benchmarks**: benchmarks covering unstructured (NarrativeQA (Kočiský et al., 2018), GovReport (Huang et al., 2021)) and structured (RepoBench (Liu et al., 2023), LCC (Guo et al., 2023)) text, evaluating zero-shot transfer against SOTA compressors. **(3) Financial-250k**: a proprietary industrial

Table 1: Zero-shot transfer from Financial-250k to LongBench against representative compressors. All compressed methods operate under RR $< 15\%$. **Bold**: best among compression methods.

Category	Method	HotpotQA		QMSum		Qasper	
		RR	F1	RR	ROUGE-L	RR	F1
Reference	Full Context	100%	0.546	100%	0.151	100%	0.420
Heuristic	Latest Truncation	12.7%	0.248	13.4%	0.043	12.4%	0.204
Prompt Compression	SelectiveContext	14.4%	0.215	14.4%	0.063	14.9%	0.215
	LLMLingua2	8.9%	0.395	12.6%	0.132	12.1%	0.352
	LongLLMLingua	11.4%	0.494	11.0%	0.147	9.4%	0.387
Retrieval -Based	Dense Top-K	9.5%	0.443	9.9%	0.095	9.4%	0.315
	Sparse (TF-IDF)	10.9%	0.401	10.8%	0.086	10.4%	0.341
Ours	**TaCo**	11.1%	**0.531**	10.8%	**0.149**	10.1%	**0.405**

benchmark (Appendix D) with documents up to 3.92M characters, 44.62% null rate, and structural heterogeneity (table, JSON, text), evaluating compression performance under extreme conditions.

Solver. All experiments use GPT-4.1 as the solver $\mathcal{M}$ with greedy decoding (temperature $= 0$) to ensure deterministic utility evaluation. For academic benchmarks, the full-context baseline uses the complete untruncated input within the solver's context window; for Financial-250k, it is instantiated via sliding-window execution with 10% overlap due to context limits, representing the maximally feasible complete-input approximation. Accordingly, all Less-is-More conclusions are interpreted relative to the respective executable setting for each dataset.

Baselines. We compare against three tiers. **(1) Heuristics**: Latest Truncation, Rule-Based Strategy. **(2) Prompt Compression**: SelectiveContext (Li et al., 2023), LLMLingua2 (Pan et al., 2024), LongLLMLingua (Jiang et al., 2024), TPC (Cao et al., 2024), Adaptive QuerySelect (Liskavets et al., 2025). **(3) Retrieval-based**: Dense Top-K (BGE-M3 (Chen et al., 2024)), Sparse (TF-IDF), and BGE-M3 + Schema + MMR—a hybrid baseline combining dense retrieval, schema-aware filtering, and diversity maximization.

Evaluation Protocol. For academic benchmarks, we report F1 (QA), ROUGE-L (summarization), and Edit Similarity (code) under RR $< 15\%$. For Financial-250k, we additionally report Score $= 0.35(1 - \text{RR}) + 0.40\,\mathcal{S}_{\text{consist}} + 0.25\,\text{F1}$, consistent with Eq. (3). F1 thresholds are set at the PR-BEP, a threshold-agnostic protocol requiring no per-method calibration and ensuring fair comparison. All results are averaged over 5 runs (std $< 1.0\%$).

4.2 RQ1: Zero-Shot Transfer Across Domains

RQ1 asks whether TaCo learns solver-level reasoning preferences that transfer zero-shot across domains. We test this at two levels: *breadth* via representative baselines on LongBench (Table 1), and *depth* via contemporary SOTAs on four native long-context benchmarks (Table 2).

Breadth: Zero-Shot Transfer to LongBench. The policy is trained exclusively on Financial-250k and evaluated zero-shot on LongBench. TaCo consistently outperforms all representative baselines: F1 $= 0.531$ on HotpotQA, ROUGE-L $= 0.149$ on QMSum, and F1 $= 0.405$ on Qasper, surpassing the strongest baseline LongLLMLingua by 3.7% / 0.2% / 1.8% absolute.

Depth: Contemporary SOTAs on Native Long-Context Benchmarks. (1) On structured tasks (RepoBench, LCC), contemporary point-wise compressors suffer catastrophic Structure Blindness: TPC drops to $50.4\%/58.6\%$ and Adaptive QuerySelect collapses to $40.9\%/45.7\%$, while TaCo achieves **72.5% / 62.2%**—surpassing the full-context baseline ($70.2\%/60.3\%$). This Less-is-More effect confirms that solver-aligned distillation actively filters distractors beyond what full context achieves. The heuristic set-wise pipeline (BGE-M3+Schema+MMR) still lags TaCo by $9.0\%/9.4\%$ absolute, demonstrating that heuristic diversity cannot substitute for solver-in-the-loop RL. **(2) On unstructured tasks** (NarrativeQA, GovReport), TaCo outperforms all compressors ($28.3\%/30.4\%$) but does not surpass full-context (31.5% / 32.9%)—an interpretable boundary condition: these unstructured corpora carry mild formatting cues (dialogue layout in scripts, Markdown headings in

Table 2: Zero-shot transfer from Financial-250k to four native long-context benchmarks against contemporary SOTAs. Crucially, Full Context ingests the prompt natively without sliding windows. All compressed methods operate under RR $< 15\%$. **Bold**: best among compression methods.

Category	Method	NarrativeQA	GovReport	RepoBench	LCC
Reference	Full Context	31.5%	32.9%	70.2%	60.3%
Prompt Compression	TPC	26.5%	27.2%	50.4%	58.6%
	Adaptive QuerySelect	25.5%	23.5%	40.9%	45.7%
Retrieval -Based	Dense Top-K	22.1%	23.6%	56.4%	48.2%
	BGE-M3+Schema+MMR	24.3%	26.8%	63.5%	52.8%
Ours	**TaCo**	**28.3%**	**30.4%**	**72.5%**	**62.2%**

Table 3: Performance on Financial-250k dataset. Score $= 0.35 \cdot (1 - \text{RR}) + 0.40 \cdot \mathcal{S}_{\text{consist}} + 0.25 \cdot \text{F1}$. All compressed methods operate under RR $< 15\%$. **Bold**: best among compression methods.

Category	Method	RR	Consistency	F1	Score
Reference	Full Context	100.00%	0.8999	0.8364	0.5691
Heuristic	Rule-Based Strategy	9.83%	0.8141	0.8316	0.8491
Prompt Compression	SelectiveContext	11.74%	0.6009	0.8365	0.7584
	LLMLingua2	11.10%	0.6150	0.8411	0.7674
	LongLLMLingua	9.63%	0.6814	0.8119	0.7918
Retrieval-Based	Dense Top-K	9.60%	0.7662	0.8360	0.8319
	Sparse (TF-IDF)	8.17%	0.6672	0.7746	0.7819
Ours	**TaCo Extreme**	**2.18%**	0.7797	0.8194	0.8593
	TaCo Balance	9.71%	**0.8277**	**0.8539**	**0.8606**

reports) that the symbolic module partially captures, but the structural prior is less decisive than on structured data, so the Less-is-More denoising effect is inherently weaker.

4.3 RQ2: Extreme Industrial Compression

RQ2 asks whether TaCo maintains compression efficiency and task effectiveness under extreme industrial conditions. Financial-250k presents three concurrent challenges: documents up to **3.92M characters**, a **44.62% null rate**, and **structural heterogeneity** (table, JSON, text). We report two operating points: *TaCo Extreme* (RR$= 2.18\%$) and *TaCo Balance* (RR$= 9.71\%$).

Results (Table 3). TaCo Extreme retains 98% of full-context F1 (0.8194 vs. 0.8364) at only 2.18% token retention. TaCo Balance **surpasses** full-context F1 (0.8539 vs. 0.8364) at 9.71% token retention—better task performance at $10\times$ lower token cost. The Less-is-More effect here is consistent with Table 2: in high-noise mixed-structured long contexts, the full context degrades solver reasoning by flooding attention with irrelevant records, and solver-aligned distillation acts as a denoising filter.

Baseline failures map directly to the taxonomy in Sec. 3.1: Dense Top-K (F1 0.8360) suffers Semantic Collapse, retrieving semantically similar but redundant records while missing complementary evidence; LLMLingua2 (F1 0.8411) suffers Structure Blindness, as its token-level scoring is agnostic to structural boundaries, corrupting the integrity of heterogeneous structured records. The strongest non-learned baseline, Rule-Based Strategy (F1 0.8316) encodes domain knowledge through handcrafted heuristics, yet its fixed rules cannot generalize beyond the patterns seen at design time, and still falls below TaCo Balance.

4.4 RQ3: Ablation Study With Failure Mode Verification

RQ3 asks whether each module mitigates its corresponding failure mode. We answer at two levels: *ablation study* via component removal on academic benchmarks (Task Score in Table 4) and

Table 4: Ablation study and mechanistic diagnostics on academic benchmarks, MMR variants instantiate *w/o RL Policy*. Each cell reports *CER (%) / PCD / Task Score (%)*. **Bold**: best per column.

Method	NarrativeQA	GovReport	RepoBench	LCC
TaCo (Full)	**78.4** / 0.45 / **28.3**	**86.5** / 0.48 / **30.4**	**95.7** / 0.44 / **72.5**	**85.8** / 0.41 / **62.2**
w/o Symbolic	55.8 / 0.43 / 22.1	59.1 / 0.46 / 23.6	64.6 / 0.41 / 56.4	61.3 / 0.38 / 48.2
w/ Regex Extractor	61.2 / 0.44 / 24.2	66.8 / 0.47 / 26.5	76.5 / 0.43 / 62.8	72.4 / 0.40 / 54.6
w/o Manifold	78.0 / 0.22 / 24.5	86.0 / 0.25 / 26.2	95.2 / 0.18 / 61.2	85.2 / 0.15 / 51.5
w/ TF-IDF Manifold	78.1 / 0.27 / 23.0	86.2 / 0.34 / 27.3	95.5 / 0.37 / 68.1	85.4 / 0.32 / 58.4
MMR (F1-Max)	56.2 / 0.38 / 24.3	61.4 / 0.40 / 26.8	68.5 / 0.35 / 63.5	64.1 / 0.33 / 52.8
MMR (Div-Max)	46.5 / **0.52** / 19.8	49.2 / **0.55** / 21.4	54.3 / **0.49** / 52.1	50.7 / **0.46** / 44.5

Table 5: Ablation Study on Financial-250k dataset, MMR (F1-Max) instantiate *w/o RL Policy*. Score $= 0.35 \cdot (1 - \text{RR}) + 0.40 \cdot \mathcal{S}_{\text{consist}} + 0.25 \cdot \text{F1}$. **Bold**: best per column.

Variant	RR	Consistency	F1	Score
TaCo (Full)	**2.18%**	**0.779**	**0.819**	**0.859**
w/o Symbolic	6.79%	0.766	0.802	0.833
w/o Manifold	2.15%	0.675	0.739	0.797
w/o RL Policy	4.12%	0.690	0.800	0.811

Financial-250k (Table 5), and *mechanistic diagnostics* via CER/PCD probes on academic benchmarks (Table 4) and failure-mode case studies on Financial-250k (Appendix F). Each ablation variant is designed to re-expose exactly one failure mode from Sec. 3.1.

Ablation Study. (1) w/o Symbolic: on academic benchmarks, Task Score drops 6.2–16.1% absolute, with the largest degradation on code tasks (RepoBench: $72.5\% \rightarrow 56.4\%$); on Financial-250k, RR spikes $3.1\times$ ($2.18\% \rightarrow 6.79\%$) while F1 remains comparable, indicating that the policy compensates by retaining far more tokens. **(2) w/o Manifold**: on academic benchmarks, Task Score drops uniformly across all four datasets; on Financial-250k, F1 suffers the largest drop ($0.819 \rightarrow 0.739$, -9.8%) with Consistency collapsing ($0.779 \rightarrow 0.675$). **(3) w/o RL Policy** (instantiated as MMR variants with $[m_i, s_i, g_i]$): on academic benchmarks, both MMR variants (Div-Max and F1-Max) fail to match TaCo's task score across all datasets, regardless of the trade-off between diversity and relevance; on Financial-250k, Consistency collapses -11.4% while F1 drops -2.3%, suggesting that RL Policy contributes to both task performance and output consistency.

The three signatures are asymmetric and non-overlapping, suggesting that each variant targets a distinct failure source. We verify this interpretation mechanistically below.

Mechanistic Diagnostics. To attribute each degradation signature to its corresponding failure mode, we apply two probes to academic benchmarks (Table 4) and failure-mode case studies to Financial-250k (Appendix F). **Critical Entity Retention (CER)** measures exact-match recall of task-specific semantic anchors (e.g., bridge names, numerical values, exact timestamps for NarrativeQA/GovReport, class/function names, API signatures, cross-file variables for RepoBench/LCC), targeting Structure Blindness. **Pairwise Cosine Distance (PCD)** measures average pairwise cosine distance among selected units, targeting Semantic Collapse.

The results reveal a clean dissociation confirming module orthogonality. **(1) *w/o Symbolic* re-exposes Structure Blindness**: CER collapses by 23–31% absolute across all datasets, while PCD remains unaffected, confirming that the Symbolic module is the primary driver of structural integrity. The gap is widest on code tasks ($95.7\% \rightarrow 64.6\%$), where discrete API boundaries are most critical. Crucially, replacing the symbolic module with a Regex Extractor also suffers a severe CER drop, confirming that rule-based extraction is fundamentally more brittle than schema-guided extraction. **(2) *w/o Manifold* re-exposes Semantic Collapse**: PCD drops 0.23–0.26 absolute across all datasets, while CER remains unaffected, confirming that the Manifold module is the primary driver of selection diversity. Crucially, replacing the BGE-M3 embeddings with sparse lexical features (TF-IDF) also fails to recover the task score and PCD, confirming that mitigating Semantic Collapse explicitly requires continuous semantic geometry, as mere lexical clustering cannot identify paraphrased redundancy. **(3) *w/o RL Policy* re-exposes Solver Mismatch**: MMR (F1-Max) maximizes relevance

at the cost of diversity, while MMR (Div-Max) maximizes diversity at the cost of relevance, yet both fall short of TaCo, confirming that the RL module is the primary driver of solver mismatch.

No single ablation recovers CER, PCD, and task score simultaneously—only the full system does. The case studies in Appendix F provide complementary qualitative evidence on Financial-250k, where each failure mode is directly observable without TaCo's solution.

5 LIMITATIONS

Schema Dependency of Symbolic State Extraction. The Symbolic module assumes a *recoverable* schema structure. For fully unstructured free-text documents such as HotpotQA (Appendix E.2), P degenerates to a sentence-boundary extractor and s_i provides no structural prior beyond m_i. We view this as a deliberate design choice: TaCo exhibits graceful degradation on HotpotQA/QMSum/Qasper (Table 1), and extending to schema-free corpora via unsupervised structure induction remains an open direction.

Heuristic Nature of Geometric Manifold Coverage. The Manifold module serves as a geometry-induced regularizer rather than a theoretically optimal diversity estimator. A residual failure mode arises when units that stand in logical entailment relations are nonetheless mapped to proximate regions of $\mathcal{Z}$, leading the Manifold module to misclassify them as redundant and placing entailment-critical units at risk of being discarded (see Appendix F.2). We regard this as a bounded limitation: our objective is sufficient inductive bias to preclude systematic failure—not exact optimality—a goal the PCD results in Table 4 directly support. Addressing this residual failure mode via entailment-aware manifold regularization remains an open direction.

Solver Specificity and Cross-Solver Generalization. TaCo's policy is trained against a fixed solver $\mathcal{M}$ and may overfit to its specific context utilization patterns; transfer degrades when the target solver exhibits substantially different context utilization behaviors (Appendix E.3). Multi-solver training is a promising mitigation left to future work.

Offline Training Amortization and Deployment Scope. TaCo follows a **train-once, infer-frequently** paradigm with an upfront cost of approximately $5000 and 30 hours on a single A100, rapidly recouped via $\sim$10$\times$ context reduction in production. TaCo is less suited for situations where data distribution shifts frequently or schema definitions evolve rapidly.

6 CONCLUSION

We presented TaCo, a solver-aligned, set-wise context distillation framework built on a diagnosis: the root cause of long-context compression failure is a mismatch between the level at which utility is optimized (point-wise or heuristic set-wise selection) and the level at which it is consumed (solver-aligned set-wise selection). TaCo targets this mismatch via a solver-in-the-loop RL framework with two complementary inductive biases—structural and geometric prior—that make intra-unit integrity and inter-unit redundancy jointly observable to the RL policy, transforming an intractable combinatorial selection problem into a stable optimization.

Empirically, TaCo achieves consistent SOTA performance among evaluated compression methods across eight benchmarks. On structured and mixed-structured tasks, the distilled subset strictly outperforms the full context (*Less-is-More*); on unstructured tasks, TaCo degrades gracefully, outperforming all compressors while approaching the full-context ceiling. Ablation study and mechanistic diagnostics confirm that the three modules are orthogonally necessary—Symbolic removal collapses CER, Manifold removal collapses PCD, Policy removal re-exposes Solver Mismatch—and policies trained on private industrial data exhibit strong zero-shot transfer to open-domain benchmarks, indicating that TaCo captures solver-specific reasoning preferences rather than dataset heuristics.

Beyond this instantiation, TaCo establishes a unifying lens for context distillation: the central challenge is *representational adequacy* of the policy state, not compression ratio or retrieval precision. Future work includes adaptive structural grounding for schema-free corpora, tighter theoretical connections to submodular optimization, multi-solver training, and surrogate reward formulations for lightweight deployment.

AI USE STATEMENT

In this work, we used generative AI models as fixed components of the proposed research pipeline. Specifically, GPT-4.1 was used as a frozen downstream solver to provide task-level feedback for reinforcement learning policy optimization and to evaluate the utility of selected contexts on both academic and industrial benchmarks. To make this evaluation deterministic, we used greedy decoding with temperature set to zero.

We also used CodeLlama-7B-Instruct as a frozen symbolic state extractor. It was used to derive schema-level structural signals from context units, including tabular, JSON, and textual inputs, which are then supplied to the policy as structural features. Neither GPT-4.1 nor CodeLlama-7B-Instruct was fine-tuned in this work.

All experimental configurations, evaluation protocols, reported results, and scientific claims were designed, reviewed, and validated by the authors. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the assistance of generative AI tools.

ETHICS STATEMENT

Data privacy and de-identification. Financial-250k was processed in a strictly secure offline environment. All personally identifiable information (PII)— names, identifiers, and precise geolocations—was irreversibly masked or hashed prior to research access. Examples presented in this paper use synthetic placeholder values (e.g., `[CITY_A]`, `[ID_MASK]`) to prevent any potential leakage.

Deployment considerations. The learned policy exhibits a utility bias toward numerical precision and factual density, a direct consequence of reward alignment with financial risk assessment tasks. We caution against out-of-the-box deployment for tasks requiring high narrative sensitivity (e.g., creative writing assistance) without domain-specific fine-tuning, as the policy may aggressively prune stylistic content assigned low utility by the financial-domain reward signal.

REPRODUCIBILITY STATEMENT

Reproducibility. Due to corporate data governance policy, code and pretrained weights can only be released upon acceptance; this timing constraint is independent of the authors' preference. Upon acceptance, we will additionally release a synthetic benchmark mirroring the statistical properties of Financial-250k—including extreme length variance, high noise ratio, and heterogeneous structure—as a public testbed for structural noise robustness without compromising privacy.

Reference Implementation for Review. To facilitate verifiability during the double-blind review, we have provided an anonymous reference implementation at `https://anonymous.4open.science/r/TaCo-B582`. This repository includes the complete RL policy architecture (reference_model.py), the RL optimization loop and inference scripts (reference_train.py, reference_inference.py), and a high-fidelity synthetic dataset containing 700 raw instances (100 for each of the 7 modules) for structural inspection.

REFERENCES

Yushi Bai, Xin Lv, Jiajie Zhang, Hongchang Lyu, Jiayi Tang, Zhidian Huang, Zhengxiao Du, Xiao Liu, Aohan Zeng, Lei Hou, Yuxiao Dong, Jie Tang, and Juanzi Li. Longbench: A bilingual, multitask benchmark for long context understanding. *arXiv preprint arXiv:2308.14508*, 2024.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*, 2020.

Zhiwei Cao, Qian Cao, Yu Lu, Ningxin Yin, Linhai Liang, Yasheng Li, Xin Jiang, Qun Liu, Xiaodan Qu, and Daojian Song. Retaining key information under high compression ratios: Query-guided compressor for LLMs. *arXiv preprint arXiv:2406.11257*, 2024.

Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 335–336, 1998.

Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. BGE M3-Embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2309.07597*, 2024.

Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating long sequences with sparse transformers. *arXiv preprint arXiv:1904.10509*, 2019.

Daya Guo, Canwen Xu, Nan Duan, Jian Yin, and Julian McAuley. Longcoder: A long-range pre-trained language model for code completion. In *Proceedings of ICML*, 2023.

Zhexin Hu, Li Wang, Xiaohan Wang, Jiajun Chai, Xiaojun Guo, Wei Lin, and Guojun Yin. ZipRL: Adaptive multi-turn context compression with hindsight response replay. *arXiv preprint arXiv:2605.28069*, 2026.

Luyang Huang, Shuyang Cao, Nikolaus Parulian, Heng Ji, and Lu Wang. Efficient attentions for long document summarization. In *Proceedings of NAACL*, 2021.

Huiqiang Jiang, Qianhui Wu, Chin-Yew Lin, Yuqing Yang, and Lili Lam. LLMLingua: Compressing prompts for accelerated inference of large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023a.

Huiqiang Jiang, Qianhui Wu, Xufang Luo, Dongsheng Li, Chin-Yew Lin, Yuqing Yang, and Lili Lam. LongLLMLingua: Accelerating and enhancing LLMs in long context scenarios via prompt compression. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024.

Jinhao Jiang, Kun Zhou, Zican Dong, Keming Ye, Wayne Xin Zhao, and Ji-Rong Wen. StructGPT: A general framework for large language model to reason over structured data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023b.

Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, and Edward Grefenstette. The NarrativeQA Reading Comprehension Challenge. *Transactions of the Association for Computational Linguistics*, 6:317–328, 2018. doi: 10.1162/tacl_a_00023.

Alex Kulesza and Ben Taskar. *Determinantal Point Processes for Machine Learning*. Now Publishers, 2012.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.

Shaoang Li, Yanhang Shi, Yufei Li, Mingfu Liang, Xiaohan Wei, Yunchen Pu, Fei Tian, Chonglin Sun, Frank Shyu, Luke Simon, Sandeep Pandey, Xi Liu, and Jian Li. Learning evidence highlighting for frozen LLMs. *arXiv preprint arXiv:2604.22565*, 2026.

Yucheng Li, Bo Dong, Frank Guerin, and Chenghua Lin. Compressing context to enhance inference efficiency of large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 6342–6353. Association for Computational Linguistics, 2023.

Barys Liskavets, Shuvendu Roy, Maxim Ushakov, Mark Klibanov, Ali Etemad, and Shane K. Luke. Prompt compression with context-aware sentence encoding for fast and improved LLM inference. *arXiv preprint arXiv:2310.03628*, 2025.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024.

Tianyang Liu, Canwen Xu, and Julian McAuley. Repobench: Benchmarking repository-level code auto-completion systems. *arXiv preprint arXiv:2306.03091*, 2023.

Baharan Mirzasoleiman, Jeff Bilmes, and Jure Leskovec. Coresets for data-efficient training of machine learning models. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.

Zhuoshi Pan, Qianhui Wu, Huiqiang Jiang, Menglin Xia, Xufang Luo, Jue Zhang, Qingwei Lin, Victor Rühle, Yuqing Yang, Chin-Yew Lin, H. Vicky Zhao, Lili Qiu, and Dongmei Zhang. LLMLingua-2: Data distillation for efficient and faithful task-agnostic prompt compression. In *Findings of the Association for Computational Linguistics: ACL 2024*, 2024.

Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*, 11:1316–1331, 2023.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, 2018.

Freda Shi, Xinyun Chen, Kanishk Misra, Nathan Scales, David Dohan, Ed Chi, Nathanael Schärli, and Denny Zhou. Large language models can be easily distracted by irrelevant context. In *Proceedings of the 40th International Conference on Machine Learning, ICML*, 2023.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2 edition, 2018. URL `http://incompleteideas.net/book/the-book-2nd.html`.

Xiaohua Wang, Zhenghua Wang, Xuan Gao, Feiran Zhang, Yixin Wu, Zhibo Xu, Tianyuan Shi, Zhengyuan Wang, Shizheng Li, Qi Qian, Ruicheng Yin, Changze Liang, Xinyu Huang, Xiaoqing Liu, and Xuanjing Huang. Searching for best practices in retrieval-augmented generation. *arXiv preprint arXiv:2407.01219*, 2024a.

Zilong Wang, Hao Zhang, Chun-Liang Li, Julian Martin Eisenschlos, Vincent Perot, Zifeng Wang, Lesly Miculivicius, Chenxi Lee, and Tomas Pfister. Chain-of-table: Evolving tables in the reasoning chain for table understanding. *arXiv preprint arXiv:2401.04398*, 2024b.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992.

Guangxuan Xiao, Yuandong Tang, Jiarui Zuo, Junxian Guo, Shang Yang, Haotian Tang, Yao Fu, and Song Han. Efficient streaming language models with attention sinks. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024.

Tianbao Xie, Chen Henry Wu, Peng Shi, Ruiqi Zhong, Torsten Scholak, Michihiro Yasunaga, Chien-Sheng Wu, Ming Zhong, Pengcheng Yin, Sida I. Wang, Victor Zhong, Bailin Wang, Chengzu Li, Connor Boyle, Ansong Ni, Ziyu Yao, Dragomir Radev, Caiming Xiong, Lingpeng Kong, Rui Zhang, Noah A. Smith, Luke Zettlemoyer, and Tao Yu. UnifiedSKG: Unifying and multi-tasking structured knowledge grounding with text-to-text language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022.

Fangyuan Xu, Weijia Shi, and Eunsol Choi. RECOMP: Improving retrieval-augmented LMs with context compression and selective augmentation. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024.

Chulhee Yun, Srinadh Bhojanapalli, Ankit Singh Rawat, Sashank J Reddi, and Sanjiv Kumar. Are transformers universal approximators of sequence-to-sequence functions? In *ICLR*, 2020.

Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, and Amr Ahmed. Big bird: Transformers for longer sequences. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020.

Zhenyu Zhang, Ying Sheng, Tianyi Zhou, Tianlong Chen, Lianmin Zheng, Ruisi Cai, Zhao Song, Yuandong Tian, Christopher Ré, Clark Barrett, Zhangyang Wang, and Beidi Chen. H_2O: Heavy-hitter oracle for efficient generative inference of large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Siyuan Zhu, Chengdong Xu, Kaiqiang Ke, and Chao Yu. Context-picker: Dynamic context selection using multi-stage reinforcement learning. *arXiv preprint arXiv:2512.14465*, 2024.

A Theoretical Analysis of Selection Paradigms

We formally establish three claims motivating solver-aligned set-wise selection: (1) point-wise methods suffer from an expressiveness gap; (2) heuristic set-wise methods suffer from a solver alignment gap; (3) solver-in-the-loop RL overcomes both and is in principle sufficient for optimizing Eq. (1).

A.1 Part I: Expressiveness Gap of point-wise Methods

Lemma A.1 (Expressiveness Gap)**.** *For any additive scoring function $f_v(\mathbf{a}) = \sum_i a_i \cdot v(x_i, q)$, there exists a task distribution $\mathbb{P}$ such that no fixed $v(\cdot, q)$ can optimize Eq. (1) in expectation over $\mathbb{P}$.*

Proof. Let $N = 4$, $k = 2$. Define two utilities:

$$\mathcal{U}^* = \mathbf{1}[\{x_1, x_2\} \subseteq \tilde{\mathcal{D}}_{\mathbf{a}}], \tag{4}$$

$$\mathcal{U}^{**} = \mathbf{1}[\{x_3, x_4\} \subseteq \tilde{\mathcal{D}}_{\mathbf{a}}]. \tag{5}$$

Optimizing $\mathcal{U}^*$ requires $v(x_1, q) \geq v(x_3, q)$ and $v(x_2, q) \geq v(x_4, q)$; optimizing $\mathcal{U}^{**}$ requires the reverse. These constraints jointly force $v(x_1, q) = v(x_3, q)$ and $v(x_2, q) = v(x_4, q)$, so f_v assigns equal scores to $\mathbf{a}^* = (1, 1, 0, 0)$ and $\mathbf{a}^{**} = (0, 0, 1, 1)$ and cannot distinguish them. Since $\mathbb{P}$ may place positive probability on both utilities, no fixed $v(\cdot, q)$ can optimize Eq. (1) in expectation. The gap is structural: f_v contains no interaction term $\phi(x_i, x_j, q)$, making it incapable of representing any utility depending on the joint composition of the selected set. □

A.2 Part II: Solver Alignment Gap of Heuristic Set-Wise Methods

Lemma A.2 (Solver Alignment Gap)**.** *For any solver-agnostic surrogate $g(\tilde{\mathcal{D}}_{\mathbf{a}}, q)$ that does not query $\mathcal{M}$, there exist two solvers $\mathcal{M}_1$, $\mathcal{M}_2$ with disjoint optimal selections such that g cannot simultaneously optimize Eq. (1) for both.*

Proof. Let $N = 4$, $k = 2$. Suppose $\mathcal{M}_1$ achieves maximum utility only on $\{x_1, x_2\}$ and $\mathcal{M}_2$ only on $\{x_3, x_4\}$. Since g does not query $\mathcal{M}$, its values are identical for $\mathcal{M}_1$ and $\mathcal{M}_2$ for all $\mathbf{a}$, so g cannot simultaneously prefer $\{x_1, x_2\}$ for $\mathcal{M}_1$ and $\{x_3, x_4\}$ for $\mathcal{M}_2$. Any fixed g must fail to optimize Eq. (1) for at least one solver. Heuristic set-wise methods (e.g., MMR, DPP) thus address the expressiveness gap of Lemma A.1 while introducing a solver alignment gap. □

A.3 Part III: Advantages of Solver-Aligned Set-Wise Selection

Theorem A.1 (Representational Sufficiency of Solver-in-the-Loop RL)**.** *A solver-in-the-loop RL policy $\pi_\theta(\mathbf{a})$ overcomes both gaps of Lemmas A.1 and A.2, and is in principle sufficient for optimizing Eq. (1).*

Proof. Overcoming the expressiveness gap. At training time, the set-level reward $R = \mathcal{U}(\mathcal{M}(\tilde{\mathcal{D}}_{\mathbf{a}}, q), y)$ propagates inter-unit interaction information into θ via:

$$\nabla_\theta \mathbb{E}_{\tau \sim \pi_\theta}[R] = \mathbb{E}\left[R \cdot \sum_{t=1}^{k} \nabla_\theta \log \pi_\theta(a^{(t)} \mid h_{i_t}^{(t)}) \right], \tag{6}$$

where $\tau = (a^{(1)}, \ldots, a^{(k)})$ denotes a full selection trajectory of k steps under sequential decoding, and i_t indexes the unit selected at step t. At inference time, the selection context vector $c^{(t)}$ in $h_i^{(t)} = [m_i \parallel s_i \parallel g_i \parallel q_{\text{emb}} \parallel c^{(t)}]$ explicitly encodes the mean embedding of all previously selected units, making $h_i^{(t)}$ directly sensitive to the current selection state at each step t. This enables π_θ to represent interaction terms $\phi(x_i, x_j, q)$ that are structurally absent from any additive scoring function f_v: the contribution of x_i is evaluated conditioned on $\mathcal{S}^{(t)}$ rather than in isolation. Since Transformer encoders are universal approximators of sequence-to-sequence functions (Yun et al., 2020), π_θ can in principle represent any such interaction term $\phi(x_i, x_j, q)$, and thereby close the expressiveness gap at both training and inference time.

Overcoming the solver alignment gap. The training objective $\max_\theta \mathbb{E}_{\mathbf{a}\sim\pi_\theta}[\mathcal{U}(\mathcal{M}(\tilde{\mathcal{D}}_\mathbf{a}, q), y)]$ uses solver output directly as the reward signal, making π_θ implicitly conditioned on $\mathcal{M}$ through the reward signal in a way no solver-agnostic surrogate g can achieve.

Tractability. Sequential decoding selects one unit per step; each step requires a single forward pass over remaining units with cost polynomial in N, and a context update of cost $O(d)$. Total training cost is $O(T)$ solver calls—one end-of-episode reward evaluation per episode—for T episodes; the k selection steps within each episode are local policy computations that require no solver access, avoiding combinatorial enumeration. Here k denotes the maximum number of selection steps per episode; in practice, selection may terminate earlier when the cumulative token count approaches κ. □

Table 6: Comparison of selection paradigms.

Paradigm	Set-level conditioning	Solver alignment	Computationally Tractable
point-wise	✗	✗	✓
Heuristic set-wise	✓	✗	✓
Solver-aligned RL	✓	✓(training solver)	✓(in principle)

B Geometric Coverage as a Proxy for Semantic Diversity

Since the ideal objective $\max I(\tilde{\mathcal{D}}; Y \mid q)$ is intractable at inference time, practical systems must rely on tractable proxies that correlate with solver utility $\mathcal{U}(\mathcal{M}(\tilde{\mathcal{D}}, q), y)$. We show that geometric dispersion in $\mathcal{Z}$ is a *necessary* such proxy under a local smoothness assumption.

B.1 Diminishing Returns and Variance Reduction

Let $U(z)$ denote the marginal utility of unit z to $\mathcal{M}$, and $\phi : \mathcal{D} \to \mathcal{Z}$ the embedding function (Sec. 3.3). We assume U is locally smooth within semantic neighborhoods:

$$|U(z_i) - U(z_j)| \;\le\; L \cdot \|\phi(z_i) - \phi(z_j)\|_2, \tag{7}$$

where L is a local Lipschitz-like constant. This applies only within local neighborhoods, not globally.

Proposition B.1 (Necessity of Coverage)**.** *Under local smoothness, any strategy allocating a non-trivial budget fraction to a single ϵ-ball in $\mathcal{Z}$ incurs diminishing marginal utility, regardless of point-wise relevance accuracy.*

Proof Sketch. Within an ϵ-ball, the marginal utility of the j-th co-selected unit is at most $L\epsilon$. Total utility is bounded by $U(z_1) + (m-1)L\epsilon$ while budget cost grows linearly in m, establishing diminishing returns. Note that this bound holds regardless of the point-wise relevance of z_1: even when $U(z_1)$ attains its maximum value, the marginal utility of each additional co-selected unit within the same ϵ-ball remains bounded by $L\epsilon$, so the diminishing-returns conclusion is independent of point-wise relevance accuracy. □

Corollary B.1 (Coverage Reduces Conditional Reward Variance)**.** *Assume $R \in [0, R_{\max}]$ almost surely, which holds since $\mathcal{S}_{\text{task}}$, $\mathcal{S}_{\text{consist}}$, and* RR *are all bounded in* $[0, 1]$*. Under local smoothness, let $h_i^{(t)} = [m_i, s_i, g_i, q_{\text{emb}}, c^{(t)}]$ (Sec. 3.4). If x_i and x_j belong to the same ϵ-ball, then*

$$\text{Var}[R \mid h_i^{(t)},\, x_j \in \tilde{\mathcal{D}}] \;\le\; \text{Var}[R \mid h_i^{(t)},\, x_j \notin \tilde{\mathcal{D}}] + 4R_{\max} L\epsilon. \tag{8}$$

When g_i encodes cluster membership and intra-cluster position, $h_i^{(t)}$ makes redundancy explicitly observable, bounding $\text{Var}[R \mid h_i^{(t)}]$ by $O(L\epsilon)$ within each cluster. Since $c^{(t)}$ and g_i encode orthogonal information, the bound holds conditioned on any realization of $c^{(t)}$.

Proof Sketch. By Proposition B.1, the reward difference between the two conditioning events is bounded by $L\epsilon$. Since $R \in [0, R_{\max}]$ almost surely, by Popoviciu's inequality $\text{Var}[R \mid A]$, $\text{Var}[R \mid B] \le R_{\max}^2/4$. Writing $\Delta = R_A - R_B$, we have $\text{Var}[R \mid A] - \text{Var}[R \mid B] = \mathbb{E}[R_A^2] - \mathbb{E}[R_B^2] - \big((\mathbb{E}[R_A])^2 - (\mathbb{E}[R_B])^2\big) \le 2R_{\max}\,\mathbb{E}[|\Delta|] + 2R_{\max}\,\mathbb{E}[|\Delta|] = 4R_{\max}\,\mathbb{E}[|\Delta|]$. By Proposition B.1, $|\Delta| \le L\epsilon$, so the variance difference is bounded by $4R_{\max}L\epsilon$, which is $O(L\epsilon)$ as claimed. Since g_i encodes cluster membership as an explicit state feature, the policy conditions on a single unified state rather than two distinct events; by the law of total variance, $\text{Var}[R \mid h_i^{(t)}] = \mathbb{E}[\text{Var}[R \mid h_i^{(t)}, \mathbf{1}_{x_j}]] \le \text{Var}[R \mid h_i^{(t)}, x_j \notin \tilde{\mathcal{D}}] + 4R_{\max}L\epsilon$, reducing conditional reward variance by $O(L\epsilon)$. □

Proposition B.1 and Corollary B.1 together establish that geometric coverage is a *necessary condition* for budget efficiency and stable credit assignment—conditions that point-wise methods cannot enforce, as their additive form renders inter-unit geometry unobservable (cf. Lemma A.1, Appendix A).

B.2 Manifold Coverage as Redundancy Control

The cluster centroids $\{\mu_1, \ldots, \mu_K\}$ minimize quantization error:

$$\sum_{x \in \mathcal{D}} \min_i \|\phi(x) - \mu_i\|_2^2, \tag{9}$$

encouraging the selected context to span a larger volume of the task-conditioned manifold and preserving distinct modes of task-relevant information jointly in $\tilde{\mathcal{D}}$.

Why Not MMR. MMR applies a uniform similarity threshold across $\mathcal{Z}$, ignoring manifold density: it over-penalizes complementary units in high-density regions and under-penalizes redundant ones in low-density regions. It therefore cannot distinguish *structurally redundant* units (same region of $\mathcal{Z}$) from *complementary* ones (distinct regions despite surface similarity). Geometric Manifold Coverage encodes each unit's manifold position as an explicit policy state feature g_i, making redundancy directly observable to the policy. Note that the solver alignment gap of heuristic methods including MMR is separately established in Lemma A.2 (Appendix A); the present discussion concerns the distinct failure mode of geometry-blindness.

Effectiveness is validated via the PCD diagnostic in Table 4 and ablation results in Table 5.

C Implementation Details

C.1 Configuration and hyperparameters

Table 7 lists the system configuration and training hyperparameters.

C.2 Implementation of Ψ

The code generation operator Ψ is realized via schema-conditioned code generation using a frozen CodeLlama-7B-Instruct model. The system prompt is applied *offline once per schema* (e.g., derived from log definitions or API documentation) rather than per sample, enforcing deterministic schema alignment and constraining the model to output strictly executable code. This design is what enables the train-once amortization discussed in Sec. 3.2.

Table 7: System configuration and training hyperparameters.

Module	Parameter	Value	Note
Symbolic State Extraction	Model	CodeLlama-7B-Instruct	Frozen
	Temperature	0	Greedy decoding
Geometric Manifold Coverage	Embedding	BAAI/bge-m3	Multi-granularity support
	Clustering	Mini-Batch K-Means	Dynamic K by length ratio
	Reference budget	$\bar{\kappa} = 0.2 \cdot l$	l is the document length
	Oversampling	$\rho = 1.5$	
RL Policy	Architecture	12-layer Transformer Encoder	Hidden dim: 768, $\sim$110M params
	Input state	$[m_i, s_i, g_i, q_{\text{emb}}, c^{(t)}]$	Feature dim: 768$\times$5
	Learning rate	1×10^{-4}	Optimizer: AdamW
	Batch size	32	Gradient accumulation: 4
	Entropy coeff	$\lambda_H = 0.01$	Prevents mode collapse
	Baseline decay	$\nu = 0.90$	EMA for variance reduction
	Reward weights	α=0.25, β=0.40, γ=0.35	Tuned on validation set
Solver	Model	GPT-4.1	Frozen
	Temperature	0	Greedy decoding

Note: *The prompt below is a generalized version; proprietary field names and business logic have been anonymized. The instruction structure and execution mechanism are identical to those used in all experiments.*

```
[INST]
Role: You are an expert Python Data Engineer specializing in log
parsing code generation.
Task: Transform a raw log segment into a structured JSON dictionary
using Python.
Instructions:
1.Input: """ {raw_log} """
2.Logic:
  - Attempt JSON parsing via json.loads; if it fails, retry with re.
  - Extract target fields: [user_id, behavior_type, amount, ...]
  - Filter out entries with null or missing target fields.
  - Apply business rules.
3.Constraints: Use only standard libraries (json, re); output
executable code only.
4.Output: Strictly print the extracted dictionary as a JSON string;
if extraction fails, print {}.
Python Code: [/INST]
```

The same prompt structure (Role/Task/Instructions/Python Code) applies to tabular inputs and code blocks; the fallback and output conventions remain identical.

C.3 RL Policy Optimization Details

Training data. The policy is trained on approximately 35k trajectories drawn from a private industrial dataset (Appendix D) comprising seven heterogeneous modules—namely consumption logs, behavioral traces, location records, search words, customer service dialogues, app usage, and credit profiles—characterized by extreme length variance, high null rates, and mixed structural formats (Appendix D, Table 10). These properties collectively instantiate all three failure modes identified in Sec. 3.1, ensuring that the learned structural and geometric priors are exercised under realistic conditions rather than narrow domain assumptions, which we attribute to the strong zero-shot transfer observed in Table 1 and Table 2.

Variance reduction. We employ an exponential moving average (EMA) of previous rewards as the baseline b_t, updated at each step:

$$b_t = \nu \cdot b_{t-1} + (1 - \nu) \cdot R_t, \tag{10}$$

where R_t is the batch-averaged reward and $\nu = 0.90$. The policy gradient uses advantage $A_t = R_t - b_t$. A learned value network or GAE would introduce an additional critic with $O(|\mathcal{D}|)$ per-step overhead; given that TaCo's efficacy derives primarily from state representation rather than policy capacity (Table 9), the lightweight EMA baseline suffices for stable convergence (Fig. 3).

Entropy regularization. To prevent premature convergence during early training, we augment the objective with an entropy bonus:

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_\theta}[R(\tau)] + \lambda \cdot \mathcal{H}(\pi_\theta), \tag{11}$$

with $\lambda = 0.01$. Gradient updates are stabilized via global gradient clipping at max norm 1.0.

Training stability. Figure 3 shows learning dynamics across 5 independent random seeds. The learning curves empirically validate our claims in Sec. 3.1: without explicit geometric and structural priors (*w/o Manifold* and *w/o Symbolic*), the policy suffers from high reward variance and sub-optimal convergence due to unstable credit assignment. In contrast, despite sparse and delayed solver feedback, TaCo strictly bounds this variance, enabling stable climbing and convergence around 1,500 steps with a final F1 standard deviation of just $\pm 0.90\%$.

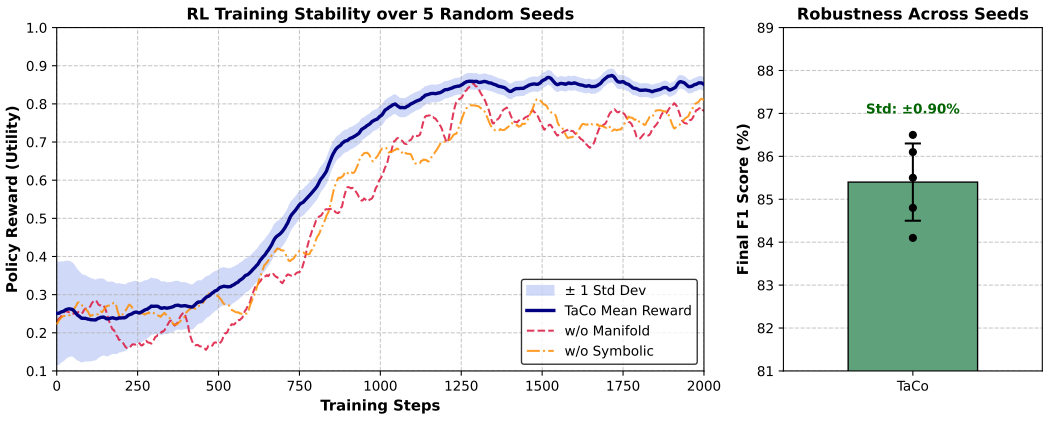


Figure 3: RL training dynamics over 5 random seeds. The shaded area denotes ± 1 std. dev. for TaCo. Notice the high reward variance and sub-optimal convergence in the *w/o Manifold* (red dashed) and *w/o Symbolic* (orange dash-dotted) ablations. This empirically demonstrates that structural and geometric priors are necessary to ensure stable credit assignment, enabling TaCo to achieve stable dominance.

Reward weight sensitivity. Table 8 reports performance on Financial-250k under reward weight variants (weights re-normalized to sum to 1.0 under each variant, and selected via grid search over $\{0.1,...,0.5\}$ in 0.05 increments on the validation split, averaged across 5 seeds). Increasing γ to 0.50 pushes the model to extreme sparsity (1.45% RR) with only minor F1 degradation, confirming the controllability of the compression–utility trade-off. Reducing β to 0.10 causes a significant F1 drop (0.7640), validating the necessity of the geometric consistency term for stable RL training—consistent with the PCD collapse observed in the ablation (Table 4).

Policy network capacity. Table 9 shows that performance is largely stable across policy capacities: the 4-layer Transformer retains 98.7% of the 12-layer Transformer's F1, confirming that TaCo's efficacy derives from the solver-aligned reward signal and state representation rather than policy capacity.

Table 8: Reward weight sensitivity on Financial-250k (average of 5 runs, std $< 1.0\%$). **Default** denotes the setting used in all main experiments. **Bold**: best per column.

Variant	$\alpha/\beta/\gamma$	RR (%)$\downarrow$	F1
Default	0.25/0.40/0.35	2.18	0.8194
High compression	0.20/0.30/**0.50**	**1.45**	0.7982
High utility	**0.50**/0.30/0.20	5.12	**0.8355**
Low consistency	0.35/**0.10**/0.55	1.95	0.7640

Table 9: Effect of policy network capacity on Financial-250k (average of 5 runs, std $< 1.0\%$). **Bold**: best per column.

Architecture	Params	F1	Consistency
Transformer (4-layer)	$\sim$30M	0.8087	0.7681
Transformer (8-layer)	$\sim$70M	0.8149	0.7743
Transformer (12-layer)	$\sim$110M	**0.8194**	**0.7797**

D FINANCIAL-250K: DATASET CHARACTERIZATION

Financial-250k is constructed from real-world industrial logs with the explicit goal of stress-testing long-context selection mechanisms under conditions that simultaneously exhibit extreme length variance, severe information sparsity, and heterogeneous structural formats. Unlike curated academic benchmarks, this dataset is designed so that standard long-context strategies (truncation, Top-K retrieval, point-wise perplexity pruning) fail *systematically* rather than due to poor tuning—each failure mode corresponding directly to one of the three structural failure modes identified in Sec. 3.1.

Privacy note. All personally identifiable information (PII)—user identifiers, phone numbers, exact addresses, transaction serial numbers—has been irreversibly anonymized or replaced with generic placeholders (e.g., `[CITY_A]`, `[ID_MASK]`). No sensitive business logic is exposed.

D.1 STATISTICAL OVERVIEW

Table 10, Figure 4 and Figure 5 summarize the seven data modules.

Table 10: Statistics of the de-identified Financial-250k dataset (50k samples). A strictly time-based split is adopted to prevent temporal leakage: 70% training, 10% validation, 20% test.

Module	Field	Null Rate (%)	Avg Len (chars)	Max Len (chars)
Consumption	`order_detail_info`	2.57	25,563	271,649
Dialogue	`talk_context`	44.62	4,328	254,554
Search	`search_word_info`	2.09	2,609	24,387
Behavior	`actions_all_info`	0.38	176,033	3,924,567
Location	`location_info`	0.62	114,402	3,739,024
Credit Profile	`pbc_detail_list`	2.33	17,572	236,119
App Usage	`app_info_array`	3.38	7,759	65,829

D.2 DATA CHALLENGE REGIMES

We characterize three representative challenge regimes.

Regime I: Extreme length variance (Behavior & Location) The `actions_all_info` and `location_info` modules dominate the tail of the length distribution, with average lengths of 176,033 and 114,402 characters respectively and individual samples reaching 3,924,567 and 3,739,024 characters—far exceeding the 128k-token context limit of typical commercial LLMs.

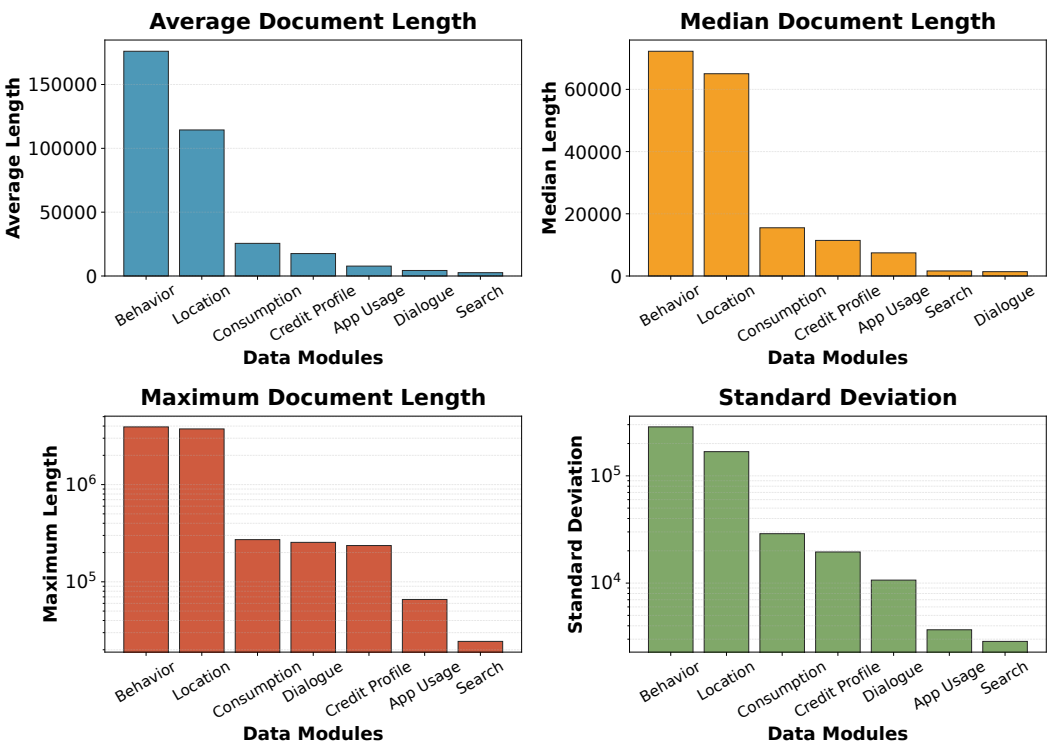


Figure 4: Financial-250k follows a heavy-tailed length distribution.

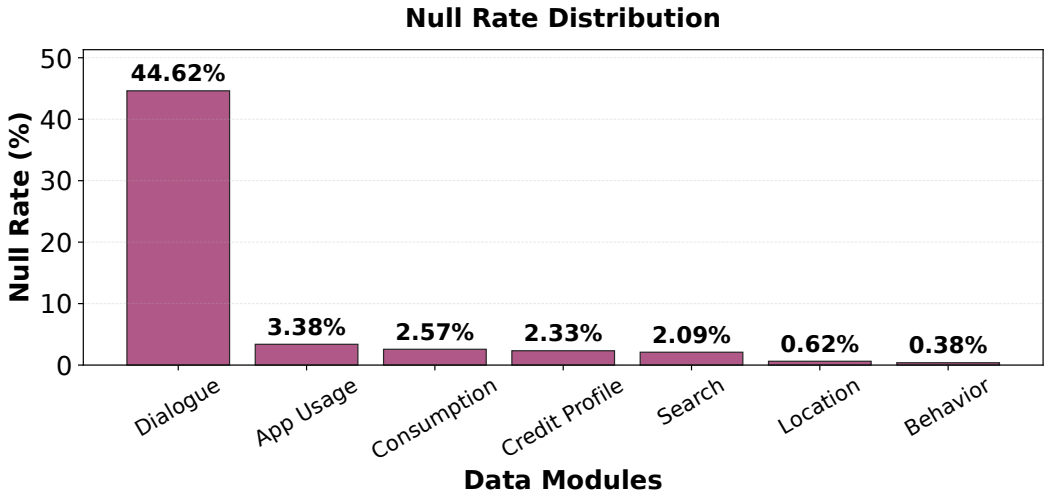


Figure 5: Financial-250k exhibits extremely low information density across modules.

This extreme length variance across modules means that a fixed-budget context window either truncates the majority of behavioral evidence or is overwhelmed by volume for the longest samples.

```
{"action_time": "2025-03-29 08:10:15", "actions_all_1d": {"type":
"exposure", "page_ident": "loan_landing_page"}}, {"action_time":
"2025-03-29 08:10:28", "actions_all_1d": {"type": "click",
"label_name": "apply_now_button"}},
// ... (thousands of similar entries) ...
```

Regime II: High noise ratio (Dialogue). The `talk_context` module exhibits the highest null rate across all seven modules at 44.62%, meaning that nearly half of all entries carry no textual content. Among the remaining non-null entries, a large fraction consists of repetitive system-generated turn pairs or single-token acknowledgements, leaving only a sparse subset of turns that carry genuine intent signals. High-frequency but low-utility events dominate point-wise embedding similarity scores, while globally decisive signals remain buried in the noise.


```
{"time": "2025-04-27 11:51:09", "scene": "service_outbound",
"text": "{"cstm": "Hello.", "agent": "Hello, I am calling
from [Platform]."}"}, {"time": "2025-04-27 11:51:10", "scene":
"service_outbound", "text": null}, {"time": "2025-04-27 11:51:12",
"scene": "service_outbound", "text": ""},
// ... (repeated empty logs) ...
```


Regime III: Heterogeneous structure (Consumption). The `order_detail_info` module illustrates heterogeneous structure at two levels. At the field level, different order categories follow distinct JSON schemas: food delivery orders populate fields such as merchant name, items, amount, and remarks, while hotel orders carry check-in and check-out dates, and electronics orders include accessory-level annotations. At the value level, the amount field exhibits categorically different distributions across order types: food delivery transactions cluster at low magnitudes, whereas hotel and electronics orders carry substantially higher values. Standard embedding methods fail to handle such heterogeneity: field-level differences across categories cause embeddings to conflate structurally distinct order types, while the amount field's magnitude semantics are invisible to token-level representations that treat "46" and "12,999" as equally arbitrary numeric strings.


```
[{"items": "[Food_Item_1], ...", "category": "Food", "merchant":
"[FastFood_Chain_A]", "amount": "46.0", "remarks": "None"},
{"check-in": "2024-11-15", "category": "Hotel", "merchant":
"[Pharmacy_B]", "check-out": "2024-11-20", "amount": "2135.0",
"city": "[City_X]"},
// ... (200+ similar small transactions) ...
{"accessory": "None", "category": "Electronics", "merchant":
"[Tech_Store_C]", "item": "High-end Laptop", "amount": "12999.0"}]
```


D.3 Prediction Task Definition

The downstream task is **binary classification over heterogeneous long-context sequences**.

- **Input:** Full heterogeneous context formed by concatenating all seven modules.
- **Output:** Binary label indicating whether the user applies for a loan within 7 days.
- **Class imbalance:** Positive class ratio $\approx 2.83\%$, creating a severe needle-in-a-haystack retrieval problem.
- **Primary difficulty:** Task-relevant evidence corresponds to a sparse subsequence of interactions embedded within a massive background of irrelevant or redundant logs, requiring precise evidence selection under extreme noise and strict token budget constraints.

E Additional Experimental Results

E.1 Latency Analysis

Table 11 presents end-to-end latency across context lengths from 0–5k to 100k+ tokens, measured on a single A100 GPU (batch size 1). We report compress time (Com), solver inference time (Inf), and total latency for GPT-4.1 on the full context, LLMLingua-2 at 10% retention, and TaCo at 10% and 2.18% retention.

Offline Precomputation vs. Online Inference. The heavy-computational tasks—code generation (symbolic module) and K-Means clustering (manifold module)—are performed offline once per

Table 11: End-to-end latency comparison across context lengths from 0–5k to 100k+ tokens.

	GPT-4.1	LLMLingua-2 (10%)			TaCo (10%)			TaCo (2.18%)		
Context	**Total**	**Com**	**Inf**	**Total**	**Com**	**Inf**	**Total**	**Com**	**Inf**	**Total**
0–5k	1.5	0.10	0.8	0.90	0.12	0.8	0.92	0.12	0.60	0.72
5–10k	3.2	0.10	0.9	1.00	0.24	0.9	1.14	0.24	0.65	0.89
10–15k	5.0	0.20	1.0	1.20	0.36	1.0	1.36	0.36	0.70	1.06
15–20k	7.0	0.25	1.2	1.45	0.48	1.2	1.68	0.48	0.75	1.23
20–25k	9.2	0.30	1.3	1.60	0.60	1.3	1.90	0.60	0.80	1.40
25–30k	11.5	0.35	1.5	1.85	0.72	1.5	2.22	0.72	0.85	1.57
30–35k	14.0	0.40	1.6	2.00	0.84	1.6	2.44	0.84	0.90	1.74
35–40k	16.8	0.50	1.8	2.30	0.96	1.8	2.76	0.96	0.95	1.91
40–45k	19.8	0.55	1.9	2.45	1.08	1.9	2.98	1.08	1.00	2.08
45–50k	23.0	0.60	2.1	2.70	1.20	2.1	3.30	1.20	1.05	2.25
50–55k	26.5	0.65	2.2	2.85	1.32	2.2	3.52	1.32	1.10	2.42
55–60k	30.2	0.70	2.4	3.10	1.44	2.4	3.84	1.44	1.15	2.59
60–65k	34.1	0.80	2.5	3.30	1.56	2.5	4.06	1.56	1.20	2.76
65–70k	38.2	0.85	2.6	3.45	1.68	2.6	4.28	1.68	1.25	2.93
70–75k	42.5	0.90	2.8	3.70	1.80	2.8	4.60	1.80	1.30	3.10
75–80k	47.0	0.95	3.0	3.95	1.92	3.0	4.92	1.92	1.35	3.27
80–85k	52.0	1.00	3.2	4.20	2.04	3.2	5.24	2.04	1.40	3.44
85–90k	57.5	1.05	3.5	4.55	2.16	3.5	5.66	2.16	1.45	3.61
90–95k	63.5	1.10	3.8	4.90	2.28	3.8	6.08	2.28	1.50	3.78
95–100k	70.0	1.15	4.1	5.25	2.40	4.1	6.50	2.40	1.55	3.95
$>$100k	78.0	1.20	4.5	5.70	2.52	4.5	7.02	2.52	1.60	4.12

data source, and amortized across all samples. The compress time for TaCo strictly measures 3 operations: symbolic code execution, geometric feature calculation, and policy inference.

Latency profile. Three trends are evident. First, GPT-4.1's native latency grows super-linearly, escalating from 1.5s at 5k tokens to 78.0s at 100k+, making real-time interaction with long contexts impractical. Second, TaCo introduces higher compress overhead than LLMLingua-2 (e.g., 2.40s vs. 1.20s at 100k), reflecting the additional cost of symbolic code execution and geometric feature calculation; this overhead is negligible relative to the reduction in solver latency. Third, TaCo (2.18%) achieves the lowest total latency across the entire context spectrum: at 100k+ tokens, total latency is 4.12s vs. 78.0s for the full-context baseline, a $\sim$19$\times$ speedup consistent with the $\sim$46$\times$ token reduction reported in Sec. 4.

E.2 Retention Heatmap Visualization

Fig. 6 presents a retention heatmap from HotpotQA (LongBench), illustrating three properties of TaCo's selection behavior that directly correspond to the theoretical claims in Sec. 3.1.

Graceful Degradation of Structural Prior. HotpotQA passages are free-form prose without recoverable schema structure; the code executor P degenerates to a sentence-boundary extractor and s_i provides no structural prior beyond m_i. Consequently, the structural prior contributes minimally to selection on this dataset, and TaCo relies primarily on the geometric prior g_i and solver feedback to drive selection. This is consistent with the graceful degradation behavior discussed in Sec. 5: on schema-free corpora, the remaining two components sustain competitive performance nonetheless.

Mitigation of Semantic Collapse. The policy suppresses passages (e.g., Passages 2, 4, 7 regarding employment, jurisdiction, and corporate law) that are structurally similar (legal case citation format) but semantically irrelevant to the specific query. Unlike Top-K retrieval, which over-samples the broad "Legal Case" cluster and saturates the budget with redundant units, the Geometric Manifold Coverage module encodes each unit's position on the semantic manifold as explicit policy state

User Question

Which case was brought to court first Miller v. California or Gates v. Collier ?

Visualization Key: ***Darker Background*** *indicates high retention probability (kept by Policy). Gray text indicates low utility tokens (discarded/skipped). The policy correctly highlights* ***Entity Names*** *and* ***Dates*** *while filtering out irrelevant text.*

Passage 1: Trusty system (prison)... The "trusty system" was a penitentiary system of discipline and security enforced in parts of the United States until the 1980s... The case **Gates v. Collier** ended the flagrant abuse... Other states using the trusty system were also forced to give it up under the ruling. Abolition **Gates v. Collier** (Gates v. Collier Prison Reform Case, **1970–1971**) ended the flagrant abuse... On October 20, 1972, Federal Judge William Keady ordered the end of racial segregation in prison residential quarters...

Passage 2: Brockmeyer v. Dun & Bradstreet... 113 Wis. 2d 561... was a case in which the Wisconsin Supreme Court first identified that Wisconsin has some judicial exceptions to the employment at will doctrine. Facts Charles J. Brockmeyer was employed at investment firm Dun & Bradstreet as a district manager of credit services... *[...Policy drops Employment Law context...]*

Passage 3: **Miller v. California**, 413 U.S. 15 (**1973**), was a landmark decision of the U.S. Supreme Court... modifying its definition of obscenity from that of "utterly without socially redeeming value" to that which lacks "serious literary, artistic, political, or scientific value". Background In **1971**, Marvin Miller... sent out a brochure advertising books... California lawmakers wrote the statute based on two previous Supreme Court obscenity cases... Decision... **Miller** was put over for reargument... and did not come down until **June 1973**...

Passage 4: Gasser v MISAT... European Court of Justice regarding the interpretation of the Brussels convention... The court's decision was considered problematic as it favoured the uniformity of application of the Brussels regime jurisdictional rules temporarily over party autonomy. Passage 5: Adams v. Burke... 84 U.S. (17 Wall.) 453 (1873)... patent monopoly... The product becomes the complete property of the purchaser and "passes without the monopoly." Passage 6: Jones v. Cunningham... writ of habeas corpus... Passage 7: Cheff v. Mathes... corporate change of control setting... Holland Furnace Company manufactured home furnaces. *[...Policy skips massive blocks of irrelevant Corporate and Patent Law...]*

Passage 8: **Gates v. Collier**, 501 F.2d 1291 (5th Cir. **1974**), was a landmark decision... that brought an end to the trusty system... It was the first case in a body of law developed in the Fifth Circuit Court of Appeals holding that a variety of forms of corporal punishment against prisoners constituted cruel and unusual punishment. District court case On February 8, **1971**, Nazareth Gates... filed a class action...

Passage 9: WN Hillas & Co Ltd v Arcos Ltd... English contract law... Passage 10: Fletcher v. Peck... Yazoo lands sales...

Figure 6: Retention heatmap on a HotpotQA (LongBench) example. Retained units are highlighted; suppressed units are dimmed.

features g_i, making inter-unit redundancy directly observable in $h_i^{(t)}$ prior to selection—the precise failure mode identified as Semantic Collapse in Sec. 3.1.

Set-Level Selection Beyond Additive Scoring. The policy identifies and retains bridge entities (e.g., timestamps 1970–1971, 1973, 1974 linked to *Gates v. Collier*), confirming that TaCo optimizes for joint information utility—preserving the inter-unit connectivity required for multi-hop reasoning—rather than independent token probability. This is the set-level selection behavior predicted by the solver-aligned objective in Eq. (1) and cannot be achieved by additive scoring functions (Lemma A.1, Appendix A).

E.3 Cross-Solver Ablation

Table 12 evaluates TaCo's compressed context across three solvers—GPT-4.1, DeepSeek-V3.1, and Gemini-2.5-Pro—under both full-context and TaCo-compressed (9.71% retention) conditions.

Consistency drops across all solvers, but unevenly. Under compression, all three solvers exhibit a decline in Consistency relative to their full-context counterparts. However, the magnitude differs substantially: GPT-4.1 drops from 0.8999 to 0.8277 (-0.0722), whereas DeepSeek-V3.1 and Gemini-2.5-Pro decline more sharply, from 0.8114 to 0.6900 (-0.1214) and from 0.9012 to 0.7779 (-0.1233) respectively. The smaller Consistency drop for GPT-4.1 indicates stronger alignment between TaCo's distilled context and GPT-4.1's context utilization patterns, consistent with GPT-4.1 serving as the fixed solver $\mathcal{M}$ during policy training.

Table 12: Cross solver ablation on Financial-250k dataset. Score $= 0.35 \cdot (1 - \text{RR}) + 0.40 \cdot \mathcal{S}_{\text{consist}} + 0.25 \cdot \text{F1}$. **Bold**: best per column.

Solver	RR	Consistency	F1	Score
GPT-4.1 (Full Context)	100.00%	0.8999	0.8364	0.5691
GPT-4.1 (TaCo's Context)	9.71%	0.8277	**0.8539**	**0.8606**
DeepSeek-V3.1 (Full Context)	100.00%	0.8114	0.8316	0.5325
DeepSeek-V3.1 (TaCo's Context)	9.71%	0.6900	0.7978	0.7915
Gemini-2.5-Pro (Full Context)	100.00%	**0.9012**	0.8325	0.5686
Gemini-2.5-Pro (TaCo's Context)	9.71%	0.7779	0.8048	0.8284

F1 reveals a less-is-more effect for the training solver. GPT-4.1 paired with TaCo's compressed context achieves a higher F1 than its full-context counterpart (0.8539 vs. 0.8364, $+0.0175$), demonstrating a clear *less-is-more* effect: noise removal via compression actively benefits task performance for the solver the policy was trained against. In contrast, DeepSeek-V3.1 and Gemini-2.5-Pro both exhibit F1 degradation under compression ($0.8316 \rightarrow 0.7978$ and $0.8325 \rightarrow 0.8048$), suggesting that the distilled context, while optimized for GPT-4.1's reasoning patterns, does not equally serve solvers with different context utilization behaviors.

Implication. Taken together, these two observations confirm that TaCo's policy may overfit to the specific context utilization patterns of the training solver, and that transfer degrades when the target solver exhibits substantially different behaviors. Multi-solver training is a promising mitigation left to future work.

F Qualitative Case Study

We analyze representative cases to concretely illustrate TaCo's working mechanism. These examples demonstrate both TaCo's strengths—empirically instantiating the three failure modes addressed in Sec. 3.1—and its limitations, consistent with the discussion in Sec. 5.

F.1 Success Case: TaCo Addresses All Three Failure Modes

Scenario. A user record spans four heterogeneous modules—browsing, search, consumption, and app usage—each with a distinct schema. Loan-relevant signals are sparse and scattered: a hotel browsing event buried among hundreds of food-and-beverage page views, a travel-related search keyword buried among dozens of promotional queries, and a recent loan-app installation buried among unrelated app events.

Input (simplified).

1. **Browsing**: 300+ `view_page` records, predominantly food-and-beverage categories; one record {`"category": "hotel", "city": "Sanya", "date": "2025-05-06"`}.
2. **Search**: 80+ queries, predominantly promotional keywords; one record {`"keyword": "Sanya resort price", "date": "2025-05-06"`}.
3. **App usage**: 200+ app events; one record {`"app_name": "[Loan_App_1]", "event": "install", "date": "2025-05-07"`}.
4. **Consumption**: routine low-signal records.

Top-K (Semantic Collapse). The loan-relevant signals—browsing, search, and app installation — each carry weak standalone relevance to the loan-intention query; their decisiveness emerges only from joint co-occurrence across modules. Meanwhile, the numerically dominant food-and-beverage and promotional records form dense clusters in the embedding space, causing greedy Top-K to repeatedly sample near-duplicate records from these clusters and fill the budget with redundant content. The sparse but complementary loan-relevant signals are crowded out entirely. The solver receives a homogeneous, high-redundancy context and predicts *No Loan Intention*.

LongLLMLingua (Structure Blindness). The token-level compressor treats all tokens uniformly across heterogeneous modules, assigning low salience scores to structural delimiters (e.g., {, }, :) that carry no standalone semantic content. As a result, critical schema boundaries are discarded mid-record, fragmenting structurally atomic units into incoherent token subsequences. The solver cannot identify the loan-app installation event from the broken context and predicts *No Loan Intention*.

BGE-M3 + Schema + MMR (Solver Mismatch). MMR balances relevance and diversity through a hand-crafted penalty term that promotes dispersion in the embedding space. However, this heuristic objective is not aligned with the solver's utility: geometric diversity in the embedding space does not correspond to the complementary cross-module signals that the solver requires. Records that are geometrically distant but query-irrelevant—such as high-volume irrelevant page browsing events and high-frequency low-amount consumption records—are retained to satisfy the coverage penalty, consuming budget that the loan-relevant signals require. The solver receives a diverse but informationally misaligned context and predicts *No Loan Intention*.

TaCo. The Symbolic State Extraction module applies module-specific schema parsers to each heterogeneous input: it identifies the `category=hotel` field in the browsing record, the travel-intent keyword in the search record, and the `event=install` field paired with a known loan-category app name, producing a compact but lossless symbolic state s_i for each unit. The Geometric Manifold Coverage module encodes the position of each unit on the semantic manifold as policy state features g_i, making the complementary relationship among the three loan-relevant signals—browsing, search, and app installation—directly observable prior to selection. The RL policy, trained against solver feedback, assigns high retention probability to precisely these three units across three distinct modules, while suppressing the redundant food-and-beverage and promotional records. The solver receives a compact, high-SNR context capturing travel intent, destination specificity, and credit-app installation, and correctly infers a *funding gap* consistent with *High Loan Intention*.

F.2 Failure Case: Geometric Proximity vs. Logical Entailment

Scenario (HotpotQA). A multi-hop query requires tracing a specific film to identify a book author, where the critical bridge sentence is geometrically proximate to a more prominent sentence in $\mathcal{Z}$.

Query. *The actor that plays Phileas Fogg in "Around the World in 80 Days" co-starred with Gary Cooper in a 1939 Goldwyn Productions film based on a novel by what author?*

Critical context.

1. (Sentence A) "Starring as a leading man in hits like *Wuthering Heights* (1939)."
2. (Sentence B) "Goldwyn used him to support Gary Cooper in the adventure tale *The Real Glory* (1939)."

TaCo selection. The Geometric Manifold Coverage module clusters sentences about "David Niven's 1939 films" into a single region C_k and identifies Sentence A as the cluster medoid $\hat{x}_k$. Sentence B lies within the ϵ-ball of Sentence A in $\mathcal{Z}$ ($\|\phi(A) - \phi(B)\|_2 < \delta$) and is flagged as geometrically redundant under extreme compression (2.18% retention). The bridge to *Gary Cooper* is dropped, and the solver fails to answer correctly.

Interpretation. This case directly instantiates the residual failure mode of Geometric Manifold Coverage identified in Sec. 5: the manifold encodes semantic distance, not logical complementarity. Sentences A and B occupy the same geometric neighborhood yet carry distinct logical entailments; the manifold-based coverage signal cannot distinguish them without explicit entailment supervision. This motivates the tighter connections to submodular optimization noted as future work in Sec. 6.