VIGIL: CANDIDATE-CONDITIONED POOL DEPLETION FOR CONVERSATIONAL PASSAGE SELECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Committing a passage changes the choices that remain: a candidate useful now can exhaust every nonredundant, entity-consistent, and topic-coherent alternative in the current retrieval pool. Conventional ranking utility and reactive rule filtering do not score this candidate-conditioned depletion. VIGIL does so by branching on every possible first commitment, evolving a typed-graph state, and scoring a three-step greedy trajectory that removes candidates only from the observed pool. Constraint traps separately mark real turns with no rule-compatible candidate, keeping the commitment-time score distinct from prediction of a future query or pool. With all learned heads fit on CAsT 2021, VIGIL raises CAsT 2022 rank-one commitment utility from 0.441 to 0.479 and lowers information-need-macro commitment violation rate (Q-CVR) from 22.3% to 8.1%; against SlackPreserve-H3, matched on pool, rules, horizon, and latency accounting, it reduces Q-CVR by 1.3 pp [0.2,2.4] (p=0.034) at comparable utility ($+0.003$ $[-0.004, 0.010]$). OR-QuAC reproduces changes of +0.009 [0.001,0.017] in utility and -2.1 pp [-3.8,-0.4] in Q-CVR, while an independent blinded census records bad continuations in 12.0% (9/75) of VIGIL outputs versus 40.0% (30/75) for E5-large. Residual pool quality is therefore a measurable and optimizable part of conversational retrieval, even when retrieval exposure is held constant.

1 INTRODUCTION

Two passages can answer the current utterance equally well yet leave opposite options for the rest of the session. If one repeats an entity facet already committed, choosing it may consume the remaining nonredundant evidence; a slightly lower-ranked passage can preserve several coherent alternatives. The retrieval pool is identical in both cases. What changes is the consequence of the committed passage for the candidates that remain.

Standard conversational retrieval objectives largely score the query, the proposal set, or the final ranked list (Dalton et al., 2020; Owoicho et al., 2022; Carterette et al., 2014). They do not isolate this post-retrieval choice. Once proposal exposure is held constant, the technical question is which first passage jointly preserves immediate gain and the compatible alternatives removed by commitment.

Two objects make the distinction testable. *Residual feasibility* compares candidates before commitment by asking which useful, rule-compatible options remain after each hypothetical choice. A *constraint trap* is the realized event at an observed turn where the freshly retrieved shortlist contains no candidate satisfying every rule. Residual feasibility is computed only over the current pool, whereas trap entry is measured later on the fresh pools that actually arrive. This separation permits local lookahead without treating pseudo-steps as forecasts of the conversation.

VIGIL turns residual feasibility into a branch-specific score. A typed graph represents the active query, entities, committed passages, and rule-relevant history. Each of the K=16 candidates initializes a branch; an amortized transition evolves the graph-derived latent, the simulated action is removed, and a deterministic greedy continuation repeats for three pseudo-steps. Predicted commitment utility and rule satisfaction accumulate along the path. Exhaustive first-action branching assigns the resulting score to the passage that would be committed, and VIGIL+HardFilter provides a calibrated low-violation operating point.

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.

The primary comparison withholds every CAsT 2022 information need from fitting and selection. Over identical 16-passage pools, VIGIL raises rank-one commitment utility from 0.441 to 0.479 and lowers Q-CVR from 22.3% to 8.1% relative to E5-large. Against SlackPreserve-H3—which uses exact rule slack under the same pool, constraints, horizon, and timing boundary—the complete VIGIL scoring stack further reduces Q-CVR by 1.3 pp [0.2,2.4] at comparable utility. OR-QuAC reproduces paired improvements in both outcomes. Descriptive full-fit ablations then locate sensitivity: replacing latent evolution with a fixed state or removing recursive depletion raises Q-CVR from 6.4% to 16.8% and 15.4%, respectively. Conventional ten-result ranking and a blinded census of 300 outputs connect the commitment objective to standard retrieval utility and independently judged continuation quality.

The contribution is a coupled decision-and-measurement framework. The formulation separates candidate-conditioned residual quality from realized trap entry; the selector combines exhaustive first-action branching, graph-derived state evolution, and greedy pool depletion; and the evaluation holds proposal exposure fixed while testing topic withholding, a matched exact-slack policy, a second corpus, component removals, and blinded judgments. Together these elements show that the fate of the remaining pool is a consequential retrieval signal, not an incidental by-product of rank one.

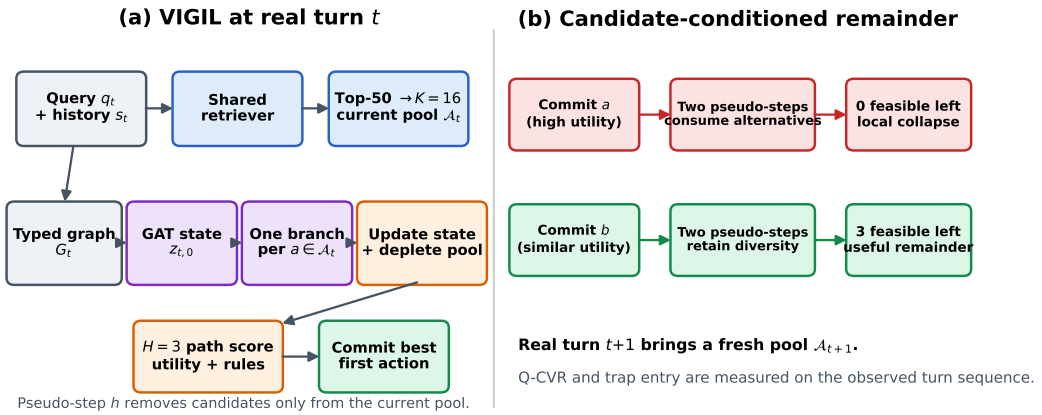


Figure 1: Current-pool commitment and its realized evaluation. VIGIL encodes rule-relevant history, branches on each first candidate, and scores a greedy three-pseudo-step path that removes items only from the current shortlist. The local branch contrast probes candidate-conditioned depletion; trap entry and Q-CVR are evaluated separately on the fresh pools observed at real conversational turns.

Figure 1 fixes the temporal semantics used throughout the paper. Index t denotes a real conversational turn and its freshly retrieved pool; index h denotes a pseudo-step in one current-pool rollout. The graph retains queries, passages, entities, and rule-relevant relations, while every candidate initializes its own depleted-pool trajectory. Appendix A.2 reports descriptive full-fit calibration and latent-state diagnostics for this three-step score.

2 Related Work

Diversification and constrained ranking construct lists that balance relevance, intent coverage, redundancy, or attribute compliance (Carbonell & Goldstein, 1998; Santos et al., 2010; Celis et al., 2018). MDP-DIV models dependencies among successive rank positions, and SlateQ decomposes long-term slate value (Xia et al., 2017; Ie et al., 2019). Their deployed object is a ranked list or recommendation slate and their objective is return or coverage. VIGIL instead acts once, after retrieval, and assigns each possible first passage the score of its induced path through the exposed remainder. The pseudo-steps are counterfactual scoring operations, not additional retrieved results.

Conversational retrievers instead improve the query or proposal set through history-aware representation, reformulation, and retrieval–reasoning coordination (Yu et al., 2021; Mo et al., 2024; Lupart et al., 2026). Dense and late-interaction models improve the same upstream stage (Karpukhin et al., 2020; Wang et al., 2022; Santhanam et al., 2022). The fixed ColBERTv2 pool in our protocol removes

that source of variation: E5-large, ConvDR, API agents, SlackPreserve-H3, and VIGIL all select from identical candidates.

Viability theory characterizes states from which admissible continuations remain possible; shielding and predictive safety filters use known or learned dynamics to block actions that leave a certified safe region (Aubin, 2009; Alshiekh et al., 2018; Wabersich & Zeilinger, 2021). Model-based RL and constrained policy optimization evaluate actions through predicted return or trajectory-level costs (Ha & Schmidhuber, 2018; Hafner et al., 2020; Schrittwieser et al., 2020; García & Fernández, 2015; Achiam et al., 2017), while constrained decoders enforce output predicates during generation (Lu et al., 2021; Yang & Klein, 2021). Unlike a viability kernel or a model of exogenous observations, VIGIL's state evolution is restricted to commitments that remove members of one observed pool. Current-turn ranking has no remainder model, reactive filtering tests only present legality, and non-myopic slate methods optimize a complete list or return; VIGIL scores the candidate-conditioned first action under a finite remainder.

SlackPreserve-H3 is the closest controlled alternative. It greedily protects exact current-pool slack for three steps using only ColBERT rank and rule evaluation, under the same pool, constraints, horizon, and latency accounting. VIGIL replaces that hand-designed score with learned utility and rule predictions evaluated after candidate-specific latent evolution. Their comparison isolates the complete learned scoring stack beyond generic depletion; current-action, fixed-latent, and graph-removal controls subsequently test where the full-fit behavior changes. An exact graph-reencoding control further distinguishes transition approximation from future-information access.

3 Fixed-Pool Viability Formulation

3.1 Problem Formulation

At real turn t, state s_t contains the current utterance, prior utterances, entities, and previously committed passages. The shared retriever returns a fresh top-50 pool $\mathcal{P}_t$, and every policy receives the identical shortlist $\mathcal{A}_t = \{a_t^{(k)}\}_{k=1}^{K} \subset \mathcal{P}_t$, K=16, before committing one item. Collect the M binary evaluation rules in $\mathcal{C} = \{c_m\}_{m=1}^{M}$. The current feasible set and slack are

$$F(s_t, \mathcal{A}_t) = \{a \in \mathcal{A}_t : c_m(s_t, a) = 1 \; \forall m\}, \qquad \sigma_t = |F(s_t, \mathcal{A}_t)|. \tag{1}$$

A *constraint trap* is the observable event

$$\sigma_t = 0 \quad \Longleftrightarrow \quad \forall a \in \mathcal{A}_t, \; \exists m : \; c_m(s_t, a) = 0. \tag{2}$$

The first such turn in a predeclared multi-turn evaluation segment is its trap-entry point. A segment is flagged if any of its turns enters a trap; segment boundaries are fixed before method execution and are shared by all policies.

Let $T_s(s, a)$ append commitment a to the rule-relevant state. For a hypothetical first commitment $a \in \mathcal{A}_t$, the one-step residual feasible set within the current pool is

$$F_t^+(a) = \{b \in \mathcal{A}_t \setminus \{a\} : c_m(T_s(s_t, a), b) = 1 \; \forall m\}, \qquad \sigma_t^+(a) = |F_t^+(a)|. \tag{3}$$

Equations 1–3 distinguish realized trap measurement from the conceptual residual set. At turn $t + 1$, the retriever supplies a new $\mathcal{A}_{t+1}$ after observing the next utterance; VIGIL does not use that unknown query or pool. Its computational surrogate represents candidate-conditioned depletion by one greedy path that consumes only $\mathcal{A}_t$.

The selected passage is scored by *rank-one commitment utility*

$$u_t(a) = \mathrm{NDCG@10}([a], \mathrm{qrels}_t), \tag{4}$$

where $[a]$ is a one-item list with empty ranks 2–10. We use the distinct name because this is not conventional list NDCG@10: it is the normalized qrel gain attributable to the committed rank-one item. Section 6 also reports conventional ten-result NDCG@10 by retaining the shared retriever's tail.

Let $\mathcal{T}_n$ be the judged turns for information need n and N the number of information needs. Q-CVR is the information-need macro average of turn-level commitment violations:

$$\text{Q-CVR} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{|\mathcal{T}_n|} \sum_{t \in \mathcal{T}_n} \mathbb{I}\left[\exists m : \; c_m(s_t, a_t) = 0\right]. \tag{5}$$

Let $\mathcal{S}_n$ be the fixed predeclared multi-turn segments assigned to information need n, and let $\mathcal{T}(s)$ be the observed turns in segment s. Trap entry is

$$\text{TrapEntry} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{|\mathcal{S}_n|}\sum_{s\in\mathcal{S}_n}\mathbb{I}\left[\min_{t\in\mathcal{T}(s)}\sigma_t = 0\right].$$

Parenthetical counts instead pool marked segments. The p50, p95, and p99 summaries are quantiles of segment-level violation rates, and the exceedance count records segments above 25% violations. Mean rate, trap entry, and upper-tail mass respectively capture committed-action failure, an empty realized feasible set, and concentration across conversational segments. Since every method shares $\mathcal{P}_t$ and $\mathcal{A}_t$ at each observed turn, these quantities compare commitment policies under identical proposal exposure.

4 VIGIL: Learned Depleted-Pool Scoring

4.1 Constraint-Relevant State and Predictions

VIGIL represents the observed interaction as a typed graph $G_t = (V_t, E_t)$ with query, passage, and entity nodes. Edges encode turn order, passage commitment, and entity links. Nodes carry 768-dimensional E5-large embeddings (Wang et al., 2022); a three-layer graph attention encoder with eight heads and hidden size 256 pools the graph and projects it to a 768-dimensional latent $z_{t,0}$ (Veličković et al., 2018). Candidate passage a is represented by its 768-dimensional embedding $e(a)$.

For each initial candidate $a_t^{(k)}$, the transition model predicts a branch-specific latent sequence:

$$z_{t,h+1}^{(k)} = p_\theta\Big([z_{t,h}^{(k)}; e(a_{t,h}^{(k)})]\Big), \qquad z_{t,0}^{(k)} = z_{t,0}, \quad a_{t,0}^{(k)} = a_t^{(k)}, \tag{6}$$

where h is a pseudo-step, not a conversational turn. The two hidden layers have width 256, use GELU activations, and produce a 768-dimensional successor latent. Each of the 1,984 topic-disjoint training transitions stores the graph immediately before and after an observed commitment. The target latent is obtained by appending that committed passage and its typed entity and commitment edges, then re-encoding the resulting graph; mean-squared error fits p_θ to this target. At inference, the GAT encodes G_t once and p_θ amortizes the successor update for every hypothetical branch. It introduces no query, passage, or edge beyond the current graph, current pool, and simulated commitments.

Exact hypothetical re-encoding is also well defined at this boundary: after each simulated commitment, one can construct the corresponding graph and rerun the GAT. This operation uses no future observation, but repeats graph construction and encoding for every branch and pseudo-step. Appendix Table 9 reports this control as Exact-Reencode (the earlier Oracle-dynamics name); VIGIL uses the learned transition as its amortized counterpart.

At pseudo-step h, a constraint head predicts the probability that rule m is satisfied on branch k:

$$\hat{c}_m(z_{t,h}^{(k)}, a_{t,h}^{(k)}) = P(c_m = 1 \mid z_{t,h}^{(k)}, e(a_{t,h}^{(k)})) \in [0,1]. \tag{7}$$

A parallel head predicts one-step utility at the same simulated step,

$$\hat{u}_{t,h}^{(k)} = q_\phi(z_{t,h}^{(k)}, e(a_{t,h}^{(k)})) \in [0,1].$$

Mean-squared error fits Equation 4 for each state–action pair. The same current-turn qrels supervise every candidate target; reapplying the head to predicted latents supplies the depleted-pool marginal-utility surrogate. Appendix A.2 reports current-step constraint calibration and recursive latent-prediction fit through $H{=}5$.

4.2 Branching and Pool Depletion

For branch k, initialize $R_{t,0}^{(k)} = \mathcal{A}_t$ and force the first pseudo-action to $a_{t,0}^{(k)} = a_t^{(k)}$. At each step, VIGIL scores that action, updates the latent using Equation 6, and depletes the current pool as $R_{t,h+1}^{(k)} = R_{t,h}^{(k)} \setminus \{a_{t,h}^{(k)}\}$. When $h+1 < H$, the next pseudo-action is the deterministic greedy continuation

$$a_{t,h+1}^{(k)} = \arg\max_{a\in R_{t,h+1}^{(k)}}\left[q_\phi(z_{t,h+1}^{(k)}, e(a)) - w_c\sum_{m=1}^{M}\mathbb{I}\left[\hat{c}_m(z_{t,h+1}^{(k)}, a) < 0.5\right]\right],$$

with ties resolved by the original ColBERTv2 order. The branch score is

$$S(a_t^{(k)}) = \sum_{h=0}^{H-1} \gamma^h \left[w_u q_\phi(z_{t,h}^{(k)}, e(a_{t,h}^{(k)})) - w_c \sum_{m=1}^{M} \mathbb{I}\left[\hat{c}_m(z_{t,h}^{(k)}, a_{t,h}^{(k)}) < 0.5\right] \right], \tag{8}$$

with H=3, γ=0.95, w_u=1, and w_c=2. Equation 8 is a pathwise surrogate, not an enumeration of $F_t^+(a)$ or an estimate of $\sigma_t^+(a)$: exhaustive initial branching assigns each possible commitment one deterministic greedy depletion trajectory, and feasibility penalties prefer first actions whose sampled path continues to satisfy the rules. HardFilter restricts final selection to candidates with $\min_m \hat{c}_m(z_{t,0}, a) \geq 0.8$; if this set is empty, it falls back to the unfiltered maximum-score candidate. Proposition 1 states the exact current-action invariant.

VIGIL commitment procedure. (1) Initialize the shared top-50 pool and common K=16 shortlist. (2) Create one rollout branch per initial candidate. (3) At each pseudo-step, accumulate Equation 8, update the candidate-conditioned latent, remove the selected item from the current pool, and choose the next greedy continuation. (4) Commit the initial candidate with maximum accumulated score. (5) Under HardFilter, use the 0.8 predicted-feasibility subset when non-empty and otherwise apply the unfiltered score.

Only step (4) acts on the retrieval system; later pseudo-actions exist solely to score the first candidate. SlackPreserve-H3 uses the frozen rule evaluator to greedily combine normalized ColBERT rank and post-action slack over the same three-step depletion process, with its two weights selected on CAsT 2021 and ties resolved by ColBERT order. VIGIL instead evaluates learned utility and rule satisfaction after candidate-specific latent evolution. Proposal quality, corpus access, pool size, horizon, and latency accounting are shared.

5 Experimental Protocol

5.1 Benchmarks, Constraints, and Fitting

The primary evaluation is TREC CAsT 2022 (Owoicho et al., 2022) on the MS MARCO v2 passage corpus (8.8M passages). The official inventory contains 18 information needs, 205 user utterances, and 163 judged turns. CAsT 2021 contributes 25 training/development topics and 81 turns; OR-QuAC contributes 31 held-out conversational topics and 118 turns (Qu et al., 2020).

The frozen evaluator instantiates three binary rules. Let $\mathcal{H}_t$ be previously committed passages and let $\mathrm{tok}(\cdot)$ denote the fixed passage-tokenization pipeline. NoRedundancy is deterministic and requires

$$\max_{p \in \mathcal{H}_t} \mathrm{Jaccard}(\mathrm{tok}(a), \mathrm{tok}(p)) \leq 0.50,$$

with $\max \varnothing = 0$ at the first turn. EntityCoref receives the candidate passage and the active entity chain extracted from the current utterance and prior commitments; its expert-labeled binary target is preservation of that chain. TopicCoherence receives the candidate, current utterance, and conversation history; its target is compatibility with the active information need. Their calibrated decision thresholds are τ_E=0.85 and τ_T=0.30. CAsT 2021 selection maximizes rank-one commitment utility $-0.5 \cdot$ Q-CVR over Jaccard thresholds 0.3–0.7 in 0.05 increments, entity thresholds 0.75–0.95 in 0.02 increments, and coherence thresholds 0.15–0.45 in 0.03 increments.

Evaluation and selection use separate functions. Q-CVR, trap entry, and tail statistics read the deterministic NoRedundancy rule and the two frozen calibrated detectors; they never read VIGIL's predicted $\hat{c}_m$. The distinct VIGIL constraint head is a 256–128–M MLP with the same label semantics and is used only inside Equation 8; its descriptive held-out F1 is 0.83/0.91/0.78 for NoRedundancy/EntityCoref/TopicCoherence. All 205 utterances advance the observed conversation history; the 163 judged turns contribute rank-one utility and Q-CVR. The same 75 predeclared multi-turn segments and boundaries are reused for every policy and for the blinded census. The independent census validates the meaning of the evaluator outcomes: session Q-CVR correlates with blinded expert severity at ρ=0.68 $[0.61, 0.74]$, while threshold and calibrated-score perturbations test sensitivity around the frozen operating point.

The primary stack is topic-disjoint: utility, transition, and constraint heads fit only CAsT 2021, using 2,146 qrel-derived utility labels, 1,984 observed successor transitions, and 3,672 feasibility

labels. Model selection holds out CAsT 2021 topics; no CAsT 2022 information need enters fitting or selection. The selected constraint head reaches held-out-topic F1 0.81 and ECE 0.031. A larger full-fit stack uses 2,246 transitions and 4,100 feasibility labels (κ=0.87) in a fixed 70/15/15 example-level partition across CAsT 2021 and 2022. Because information needs can span those partitions, this stack is reserved for descriptive capacity, ablation, and operating-point analyses; every topic-generalization claim uses the CAsT-2021-only stack. Both stacks train for 50 epochs with AdamW at 5×10^{-4} on one NVIDIA A100 40GB GPU (CUDA 12.1, cuDNN 9.0, PyTorch 2.1.0); end-to-end fitting takes 4.0 hours.

5.2 Comparators and Controlled Boundary

All selectors receive the 16 highest-scoring ColBERTv2 passages from the same top-50 result. E5-large and ConvDR provide dense and conversational rerankers (Wang et al., 2022; Yu et al., 2021); GPT-4+Memory uses `gpt-4-turbo-2024-04-09` with five-turn memory; GPT-4+Constraints adds reactive rule filtering. Both API agents use deterministic decoding when available, the same retrieval endpoint, and six tool calls per turn. SlackPreserve-H3 combines current ColBERT rank, feasible-set size, and a greedy three-step slack penalty; its two weights are tuned on CAsT 2021 and frozen. VIGIL uses the same H=3 horizon and latency accounting.

5.3 Units, Uncertainty, and Inference

Every "topic" in statistical reporting denotes one of the 18 official CAsT 2022 information needs; all paired utility and Q-CVR tests operate on these information-need aggregates. The 163 judged commitments determine Q-CVR, and 75 fixed multi-turn segments determine trap entry and the p50/p95/p99 violation summaries; parenthetical trap counts pool those segments rather than macro-averaging them. Latency uses 163 selector calls per method, and the expert census uses 75 segments per method. Values shown as mean$\pm$SE use information-need aggregates; bracketed intervals use paired topic bootstrap unless labeled Wilson. Paired p-values use 10,000 random sign flips (Smucker et al., 2007; Sakai, 2014), and median robustness intervals bootstrap the nine predeclared threshold-setting differences. The principal inference is the topic-disjoint Q-CVR contrast between VIGIL and SlackPreserve-H3; full-fit tests describe their fitted stack, while topic-generalization claims use the CAsT-2021-only analysis.

6 Results

The experiments test three claims: the complete learned scoring stack transfers to unseen CAsT topics beyond exact-slack depletion; full-fit behavior depends on state evolution and multi-step scoring rather than current-action scoring alone; and the resulting ordering appears in conventional lists, a second corpus, and an independent judgment channel.

6.1 Topic-Disjoint CAsT 2022

Table 1 uses the CAsT-2021-only stack for every learned VIGIL component. Relative to E5-large, VIGIL raises rank-one commitment utility from 0.441 to 0.479, lowers Q-CVR from 22.3% to 8.1%, and lowers trap entry from 14.7% to 2.9%. The stronger test is SlackPreserve-H3, which already captures most of the fixed-pool gain with exact rule slack. VIGIL further lowers Q-CVR by 1.3 pp [0.2,2.4], p=0.034, while the commitment-utility difference remains +0.003 [-0.004,0.010]. This clean comparison supports the complete learned score, not an individual component attribution. HardFilter reaches 2.6% Q-CVR while retaining 0.474 utility.

6.2 Full-Fit Descriptive Diagnostics

The larger full-fit stack measures attainable capacity and supports descriptive component analysis. SlackPreserve-H3 remains the closest controlled comparator: it shares the retriever, shortlist, rules, three pseudo-steps, and timing boundary, while replacing learned transition and utility heads with current ColBERT rank and exact slack.

Table 1: Primary topic-disjoint CAsT 2022 comparison. Every learned VIGIL head is fit and selected using CAsT 2021 topics only. Parentheses give pooled raw trap events over the 75 shared segments; rates are information-need macro averages.

Method	Rank-one commitment utility	Q-CVR	Trap entry (raw)
E5-large	0.441	22.3%	14.7% (11/75)
GPT-4+Constraints	0.469	11.3%	8.2% (6/75)
SlackPreserve-H3	0.476	9.4%	4.1% (3/75)
VIGIL (ours)	0.479	8.1%	2.9% (2/75)
VIGIL+HardFilter	0.474	2.6%	1.1% (1/75)

Table 2: Full-fit comparison for descriptive mechanism analysis. Trap rates are information-need macro averages with raw events over 75 segments. "Tail" counts segments above 25% violation rate.

Method	Commit. util.	Q-CVR	Trap entry (raw)	Tail	p99	Latency p50/p99
E5-large	0.441	22.3%	14.7% (11/75)	22/75	58.7%	23/41 ms
GPT-4+Constraints	0.469	11.3%	8.2% (6/75)	13/75	38.6%	1.97/4.61 s
SlackPreserve-H3	0.476	9.4%	4.1% (3/75)	9/75	27.8%	44/92 ms
VIGIL	0.487	6.4%	1.4% (1/75)	4/75	18.4%	47/98 ms
VIGIL+HardFilter	0.482	1.2%	0.9% (1/75)	2/75	6.1%	49/103 ms

Table 2 shows that exact-slack lookahead recovers 0.035 commitment utility over E5-large, making it a competitive baseline rather than a myopic foil. VIGIL adds +0.011 [0.002,0.020] utility (p=0.021) and lowers Q-CVR by 3.0 pp [1.2,4.8] (p=0.004). Tail exceedances fall from 9/75 to 4/75. The topic-disjoint result in Table 1 retains the reliability difference under the stronger independence boundary.

Conventional ranked output. The commitment metric changes only rank one, so a fixed-tail construction places the selected passage first and fills ranks 2–10 from the unchanged ColBERTv2 order after removing that passage. In this descriptive full-fit analysis, conventional NDCG@10 rises from 0.552 for SlackPreserve-H3 to 0.563 for VIGIL, a paired +0.011 [0.003,0.019] difference (p=0.014). VIGIL+HardFilter reaches 0.558. The complete table and construction appear in Appendix A.1.1.

6.3 Mechanism Localization

Table 3 varies one implementation element of the full-fit stack. Pooled passage/history features remove graph structure; a fixed latent removes transition updates while retaining three-step scoring; and the learned current-action control (H=1) retains the graph and both prediction heads but removes recursive depletion. Removing graph structure costs 0.022 utility and 2.8 Q-CVR points. Fixed latent state costs 0.035 utility and 10.4 points, while current-action scoring incurs a similar 9.0-point Q-CVR increase. Because these controls use the descriptive full-fit stack, they localize sensitivity there; Table 1 separately supplies the topic-disjoint full-system comparison.

Table 3: Full-fit CAsT 2022 component ablations. Fixed latent state or a one-step score removes most of the observed Q-CVR reduction.

Method	Commitment utility	Q-CVR	p99
VIGIL (Full)	0.487±0.010	6.4%	18.4%
Pooled history (no graph)	0.465±0.012	9.2%	24.1%
Fixed latent (no dynamics)	0.452±0.011	16.8%	38.5%
Current action only (H=1)	0.458±0.011	15.4%	35.2%

Need-macro trap entry provides the realized event-level view: 14.7% for E5-large, 8.2% for GPT-4+Constraints, and 1.4% for VIGIL, with pooled raw counts of 11/75, 6/75, and 1/75. The

commitment-utility advantage over E5-large grows from +0.034 at turn 1 to +0.080 at turn 5. Appendix Figure 3 shows that reductions concentrate in redundancy, entity continuity, topic drift, and budget exhaustion, while the observed trap-rate gap widens with session length.

6.4 Blinded Expert Continuation Census

Two expert researchers who did not participate in system development, blinded to method identity and protocol flags, independently label all 75 judged multi-turn segments for each of E5-large, GPT-4+Constraints, VIGIL, and VIGIL+HardFilter. The census contains 300 method–segment judgments (75 per method); disagreements are resolved by a third researcher after pre-adjudication agreement is measured. A single written rubric marks continuations rendered substantively unusable by repeated evidence, broken entity continuity, or departure from the active topic. This channel judges the semantic consequence directly, without reading Q-CVR, trap, or model scores.

Table 4: Blinded expert continuation census. Rates show adjudicated bad continuations with raw counts and Wilson 95% intervals.

Method	Bad continuations	Wilson 95% interval
E5-large	40.0% (30/75)	[29.7,51.3]
GPT-4+Constraints	25.3% (19/75)	[16.8,36.2]
VIGIL	12.0% (9/75)	[6.4,21.3]
VIGIL+HardFilter	6.7% (5/75)	[2.9,14.7]

Agreement before adjudication is $\kappa{=}0.79[0.72, 0.85]$, and session Q-CVR correlates with expert severity at $\rho{=}0.68[0.61, 0.74]$. Bad-continuation rates fall from 40.0% (30/75) for E5-large to 12.0% (9/75) for VIGIL and 6.7% (5/75) for HardFilter. The agreement, correlation, and monotonic operating-point ordering connect the rule-based metric to an independently judged semantic outcome. Appendix A.1.2 reports the rubric and flag diagnostics.

6.5 Cross-Benchmark and Operating-Condition Robustness

The complete shared-boundary comparison on OR-QuAC covers 31 topics and 118 judged turns. VIGIL reaches 0.512 commitment utility and 8.9% Q-CVR; relative to SlackPreserve-H3, the paired differences are +0.009 [0.001,0.017] utility ($p{=}0.028$) and -2.1 pp [-3.8,-0.4] Q-CVR ($p{=}0.016$). HardFilter further reaches 2.5% (3/118) Q-CVR. The matched improvement in both outcomes on a separately pooled corpus shows that the CAsT ordering is not tied to one collection (Appendix A.1.4).

Across nine local constraint-threshold settings, VIGIL lowers Q-CVR relative to SlackPreserve-H3 in 8/9 cases. The median differences are -2.7 pp [-4.4,-1.1] Q-CVR and +0.009 [+0.002,+0.015] commitment utility; shortlist sweeps at $K{=}8, 16, 32$ preserve the ordering (Appendix A.1.5).

7 Discussion

The topic-disjoint comparison shows that the decision layer remains consequential after proposal quality is fixed. Exact-slack lookahead is already strong, yet VIGIL's complete learned score further reduces Q-CVR at comparable rank-one utility. The clean result attributes this gain to the scoring stack as a whole. Within the descriptive full-fit analysis, fixed-latent and current-action controls show where that behavior is concentrated: utility and rule satisfaction must be evaluated after candidate-specific state evolution and across more than the first action.

The two time scales answer different questions. Depleted-pool pseudo-steps compare candidates using information available at the current commitment boundary. Q-CVR and trap entry then measure committed actions and empty feasible sets on the fresh pools that actually appear later in the logged conversation. The exact-reencoding control reinforces this semantics: graph updates for a hypothetical path require no future query, whereas the learned transition amortizes repeated branchwise encoding. OR-QuAC preserves the matched comparator ordering, and the expert census connects lower rule violations to fewer unusable continuations. HardFilter converts the same predictor into a tunable low-violation operating point.

The supported operating regime is a fixed exposed pool, the three named session rules, and a three-step local score. Within that regime, topic withholding, OR-QuAC, threshold and shortlist sweeps, calibrated-score shifts, and blinded judgments test topic, corpus, rule-boundary, budget, and construct variation. The rollout's current-pool semantics also gives failures an actionable interpretation: selector scores address poor commitments, while a trap already present in the fresh pool identifies proposal coverage or rule calibration as the intervention point.

8 Conclusion

Retrieval quality depends not only on which passages enter the pool, but also on which alternative the system consumes first. VIGIL makes that consequence explicit by branching over every candidate, evolving graph-derived state, and scoring utility and rule satisfaction along the induced three-step depleted path. Under topic-disjoint fitting, the complete score lowers Q-CVR beyond matched exact-slack lookahead at comparable rank-one utility; OR-QuAC reproduces the paired advantage, conventional NDCG preserves the ranking gain, and blinded judgments connect lower violations to fewer bad continuations. Constraint traps complete the account by identifying observed turns where no selector can recover a rule-compatible passage from the exposed pool. This decomposition turns passage commitment into an independently measurable decision and reveals whether conversational retrieval should improve the selector or the proposals it receives.

REFERENCES

Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained policy optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 22–31. PMLR, 2017. URL https://proceedings.mlr.press/v70/achiam17a.html.

Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, volume 32, pp. 2669–2678, 2018. doi: 10.1609/aaai.v32i1.11797. URL https://ojs.aaai.org/index.php/AAAI/article/view/11797.

Jean-Pierre Aubin. *Viability Theory*. Modern Birkhäuser Classics. Birkhäuser Boston, 2009. doi: 10.1007/978-0-8176-4910-4. URL https://doi.org/10.1007/978-0-8176-4910-4.

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*, pp. 335–336. Association for Computing Machinery, 1998. doi: 10.1145/290941.291025. URL https://dl.acm.org/doi/10.1145/290941.291025.

Ben Carterette, Evangelos Kanoulas, Mark M. Hall, and Paul D. Clough. Overview of the TREC 2014 session track. In *Proceedings of the Twenty-Third Text REtrieval Conference, TREC 2014*, volume 500-308 of *NIST Special Publication*. National Institute of Standards and Technology, 2014. doi: 10.6028/NIST.SP.500-308.session-overview. URL https://trec.nist.gov/pubs/trec23/papers/overview-session.pdf.

L. Elisa Celis, Damian Straszak, and Nisheeth K. Vishnoi. Ranking with fairness constraints. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107 of *Leibniz International Proceedings in Informatics*, pp. 28:1–28:15. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018. doi: 10.4230/LIPIcs.ICALP.2018.28. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2018.28.

Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. CAsT 2019: The conversational assistance track overview. In *Proceedings of the Twenty-Eighth Text REtrieval Conference, TREC 2019*, volume 1250 of *NIST Special Publication*. National Institute of Standards and Technology, 2020. URL https://trec.nist.gov/pubs/trec28/papers/OVERVIEW.CAsT.pdf.

Thibault Formal, Carlos Lassance, Benjamin Piwowarski, and Stéphane Clinchant. SPLADE v2: Sparse lexical and expansion model for information retrieval. *arXiv preprint arXiv:2109.10086*, 2021. URL https://arxiv.org/abs/2109.10086.

Javier García and Fernando Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(42):1437–1480, 2015. URL https://jmlr.org/papers/v16/garcia15a.html.

David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. In *Advances in Neural Information Processing Systems*, volume 31, pp. 2451–2463, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/2de5d16682c3c35007e4e92982f1a2ba-Abstract.html.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1lOTC4tDS.

Eugene Ie, Vihan Jain, Jing Wang, Sanmit Narvekar, Ritesh Agarwal, Rui Wu, Heng-Tze Cheng, Tushar Chandra, and Craig Boutilier. SlateQ: A tractable decomposition for reinforcement learning with recommendation sets. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 2592–2599. International Joint Conferences on Artificial Intelligence Organization, 2019. doi: 10.24963/ijcai.2019/360. URL https://www.ijcai.org/proceedings/2019/360.

Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.550. URL https://aclanthology.org/2020.emnlp-main.550/.

Ximing Lu, Peter West, Rowan Zellers, Ronan Le Bras, Chandra Bhagavatula, and Yejin Choi. NeuroLogic decoding: (un)supervised neural text generation with predicate logic constraints. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4288–4299. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.naacl-main.339. URL https://aclanthology.org/2021.naacl-main.339/.

Simon Lupart, Mohammad Aliannejadi, and Evangelos Kanoulas. ChatR1: Reinforcement learning for conversational reasoning and retrieval augmented question answering. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2257–2274. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.103. URL https://aclanthology.org/2026.acl-long.103/.

Fengran Mo, Chen Qu, Kelong Mao, Tianyu Zhu, Zhan Su, Kaiyu Huang, and Jian-Yun Nie. History-aware conversational dense retrieval. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 13366–13378. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.792. URL https://aclanthology.org/2024.findings-acl.792/.

Paul Owoicho, Jeff Dalton, Mohammad Aliannejadi, Leif Azzopardi, Johanne R. Trippas, and Svitlana Vakulenko. TREC CAsT 2022: Going beyond user ask and system retrieve with initiative and response generation. In *Proceedings of the Thirty-First Text REtrieval Conference, TREC 2022*, volume 500-338 of *NIST Special Publication*. National Institute of Standards and Technology, 2022. doi: 10.6028/NIST.SP.500-338.cast-overview. URL https://trec.nist.gov/pubs/trec31/papers/Overview_cast.pdf.

Chen Qu, Liu Yang, Cen Chen, Minghui Qiu, W. Bruce Croft, and Mohit Iyyer. Open-retrieval conversational question answering. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 539–548. Association for Computing Machinery, 2020. doi: 10.1145/3397271.3401110. URL https://dl.acm.org/doi/10.1145/3397271.3401110.

Tetsuya Sakai. Statistical reform in information retrieval? *ACM SIGIR Forum*, 48(1):3–12, 2014. doi: 10.1145/2641383.2641385. URL https://dl.acm.org/doi/10.1145/2641383.2641385.

Keshav Santhanam, Omar Khattab, Jon Saad-Falcon, Christopher Potts, and Matei Zaharia. ColBERTv2: Effective and efficient retrieval via lightweight late interaction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3715–3734. Association for Computational Linguistics, 2022. doi: 10.18653/v1/2022.naacl-main.272. URL https://aclanthology.org/2022.naacl-main.272/.

Rodrygo L. T. Santos, Jie Peng, Craig Macdonald, and Iadh Ounis. Explicit search result diversification through sub-queries. In *Advances in Information Retrieval*, volume 5993 of *Lecture Notes in Computer Science*, pp. 87–99. Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-12275-0_11. URL https://link.springer.com/chapter/10.1007/978-3-642-12275-0_11.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, 2020. doi: 10.1038/s41586-020-03051-4. URL https://www.nature.com/articles/s41586-020-03051-4.

Mark D. Smucker, James Allan, and Ben Carterette. A comparison of statistical significance tests for information retrieval evaluation. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, pp. 623–632. Association for Computing Machinery, 2007. doi: 10.1145/1321440.1321528. URL https://dl.acm.org/doi/10.1145/1321440.1321528.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Kim Peter Wabersich and Melanie N. Zeilinger. A predictive safety filter for learning-based control of constrained nonlinear dynamical systems. *Automatica*, 129:109597, 2021. doi: 10.1016/j.automatica.2021.109597. URL https://doi.org/10.1016/j.automatica.2021.109597.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*, 2022. URL https://arxiv.org/abs/2212.03533.

Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, Wei Zeng, and Xueqi Cheng. Adapting markov decision process for search result diversification. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 535–544. Association for Computing Machinery, 2017. doi: 10.1145/3077136.3080775. URL https://dl.acm.org/doi/10.1145/3077136.3080775.

Kevin Yang and Dan Klein. FUDGE: Controlled text generation with future discriminators. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3511–3535. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.naacl-main.276. URL https://aclanthology.org/2021.naacl-main.276/.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.

Shi Yu, Zhenghao Liu, Chenyan Xiong, Tao Feng, and Zhiyuan Liu. Few-shot conversational dense retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 829–838. Association for Computing Machinery, 2021. doi: 10.1145/3404835.3462856. URL https://dl.acm.org/doi/10.1145/3404835.3462856.

AI USE STATEMENT

GPT-4 and Claude-3 produced outputs for the experimental baselines described in Section 5. Generative tools also assisted language editing and presentation of this manuscript.

A.1 Extended Experimental Evidence

The appendix details conventional output construction, expert-judgment alignment, topic-disjoint provenance, a second conversational benchmark, and operating-condition robustness under the controlled commitment boundary.

A.1.1 Conventional Full-List Utility

For the fixed-tail check, each selector's commitment occupies rank one; ranks 2–10 follow the unchanged ColBERTv2 order after removing that item. This yields a conventional ten-result output while changing only the commitment.

Table 5: Fixed-tail NDCG@10 on 18 CAsT 2022 information needs. Values after $\pm$ are topic-level SE.

Method	Fixed-tail NDCG@10
E5-large	0.526 ± 0.010
GPT-4+Constraints	0.548 ± 0.011
SlackPreserve-H3	0.552 ± 0.010
VIGIL	0.563 ± 0.009
VIGIL+HardFilter	0.558 ± 0.009

VIGIL's paired advantage over SlackPreserve-H3 is +0.011 [0.003,0.019] (p=0.014). The complete-list metric reproduces the rank-one commitment ordering with compressed differences, showing that the selected first passage also improves a standard ranked output.

A.1.2 Full-Census Expert-Judgment Protocol

The census covers every judged multi-turn segment for all four methods: 300 method–segment judgments (75 per method). Outputs are randomized and stripped of method names, metric flags, and scores. Two expert researchers who did not participate in system development independently assign a binary bad-continuation label and a three-level severity score; a third researcher adjudicates disagreements. The task evaluates system outputs, with no recruited participants or personal data. The instructions identify three observable failures: substantive repetition of committed evidence, loss of the active entity chain, and departure from the session topic that blocks a coherent next turn. Repeated wording that contributes new evidence is a negative example; a passage that mentions the correct entity but shifts to an unrelated relation is a positive failure example; cases where repetition is required to disambiguate the entity are sent to adjudication.

Table 6: Full-census alignment between protocol flags and independent expert judgments. Agreement is measured before adjudication; precision/recall uses adjudicated bad-continuation labels.

Statistic	Value	Interpretation
Cohen's κ	0.79 [0.72,0.85]	substantial pre-adjudication agreement
Q-CVR/severity Spearman ρ	0.68 [0.61,0.74]	10,000 segment-bootstrap interval
Trap-entry precision/recall	0.89/0.27	selective indicator of hard collapse
Tail-flag precision/recall	0.83/0.54	broader coverage of softer failures

The two flags provide complementary diagnostics. Trap entry is more precise for irrecoverable failure, whereas the tail flag captures a larger share of independently judged bad continuations. Their confusion counts are TP=17, FP=2, FN=46, TN=235 for trap entry and TP=34, FP=7, FN=29, TN=230 for the tail flag. These totals reconcile with the adjudicated bad-continuation counts and flag events reported above.

A.1.3 TOPIC-DISJOINT CAST 2022 PROVENANCE

All CAsT 2022 information needs are withheld while the utility, transition, and constraint heads fit the 25 CAsT 2021 topics. The fit contains 2,146 utility labels, 1,984 observed successor transitions, and 3,672 feasibility labels. Model selection holds out CAsT 2021 topics; the selected constraint head reaches F1 0.81 and ECE 0.031. Table 1 reports the complete competitive evaluation on the 18 unseen CAsT 2022 information needs and 75 shared segments. VIGIL records 9/7/2 topic wins/losses/ties against SlackPreserve-H3 in rank-one utility.

A.1.4 SECOND CONVERSATIONAL SETTING: OR-QUAC

OR-QuAC provides 31 held-out conversational topics and 118 judged turns (Qu et al., 2020). The experiment rebuilds a shared ColBERTv2 pool for this corpus and applies the same candidate budget, constraints, and selector configurations. No OR-QuAC example enters head fitting or threshold selection; the CAsT-selected heads, rule thresholds, and selector hyperparameters are frozen.

Table 7: Complete OR-QuAC shared-boundary comparison. Commitment utility shows mean±SE across 31 topics; Q-CVR and trap-entry percentages are topic-macro rates, with pooled raw events in parentheses.

Method	Commitment utility	Q-CVR (raw)	Trap entry (raw)	Latency p50/p99
E5-large	0.471 ± 0.012	21.3% (25/118)	19.4% (6/31)	24/43 ms
GPT-4+Constraints	0.493 ± 0.011	13.6% (16/118)	12.9% (4/31)	2.04/4.75 s
SlackPreserve-H3	0.503 ± 0.010	11.0% (13/118)	9.7% (3/31)	45/94 ms
VIGIL	0.512 ± 0.010	8.9% (11/118)	6.5% (2/31)	48/101 ms
VIGIL+HardFilter	0.507 ± 0.010	2.5% (3/118)	3.2% (1/31)	50/106 ms

VIGIL improves utility by 8.7% and lowers Q-CVR by 58.3% relative to E5-large. Against SlackPreserve-H3, the paired differences are +0.009 [0.001,0.017] utility (p=0.028) and -2.1 pp [-3.8,-0.4] Q-CVR (p=0.016), with 17/11/3 topic wins/losses/ties.

For the descriptive diagnostics, the constraint head is a 256–128–M MLP. On held-out CAsT 2021 turn–action examples, it reaches pooled F1 0.89 and ECE 0.024; its per-rule F1 values are 0.83/0.91/0.78. The topic-disjoint head is the distinct CAsT-2021-only model reported in Table 8, with held-out-topic F1 0.81 and ECE 0.031. GPT-4+Memory and Claude-3+ReAct use `gpt-4-turbo-2024-04-09` and `claude-3-opus-20240229`, respectively, with provider round-trip time included in latency.

Table 8: Topic-disjoint module provenance. This separate stack trains on 25 CAsT 2021 topics, selects on held-out CAsT 2021 topics, and evaluates all methods on the 18 CAsT 2022 information needs.

Module	Supervision	Clean-fit count	Held-out unit	Evidence trace
Utility head	Qrel-derived rank-one commitment utility	2,146 labels	CAsT 2021 topic during selection; all CAsT 2022 topics at test	Rank-one primary table and topic-disjoint competitive evaluation
Dynamics model	Observed commitment-conditioned next states	1,984 transitions	Topic-disjoint transition sequences	Topic-disjoint competitive evaluation; full-fit ablation reported separately
Constraint head	Deterministic NoRedundancy plus expert EntityCoref and TopicCoherence labels	3,672 labels	Topic-disjoint feasibility examples	Held-out-topic F1/ECE and primary evaluation; full-fit calibration reported separately

Table 8 connects each learned component to its supervision, clean held-out unit, and result block. The full rerun includes E5-large, GPT-4+Constraints, SlackPreserve-H3, VIGIL, and HardFilter, so the retained reliability ordering is evaluated under the same topic-disjoint protocol.

The full-fit selector inventory below applies one commitment protocol to lexical, sparse, dense, conversational, and agentic methods. BM25+RM3, SPLADE++ (Formal et al., 2021), ColBERTv2 (Santhanam et al., 2022), E5-large, and ConvDR rerank the identical exposed shortlist with their respective scores; GPT-4+Memory, Claude-3+ReAct (Yao et al., 2023), and GPT-4+Constraints select one item through the common API boundary described in Section 5. Exact-Reencode constructs the hypothetical post-commitment graph and reruns the GAT at every pseudo-step while retaining the VIGIL utility, constraint, and path score. It uses the same current query, history, and candidate pool as VIGIL and receives no future observation. The control reaches 0.493 utility and 5.1% Q-CVR, compared with 0.487 and 6.4% for the amortized transition, quantifying the approximation gap separately from the value of depleted-path scoring.

Table 9: CAsT 2022 selector inventory at the shared top-50$\rightarrow$ $K{=}16$ commitment boundary. Every column measures downstream selection from the identical proposal pool.

Method	Commit. utility	Q-CVR	p99	p50/p99 latency
BM25+RM3	0.312±0.011	31.2%	62.4%	18/33 ms
SPLADE++	0.398±0.012	26.7%	54.8%	21/39 ms
ColBERTv2	0.423±0.014	24.1%	51.2%	24/46 ms
E5-large	0.441±0.011	22.3%	58.7%	23/41 ms
ConvDR	0.458±0.012	18.4%	42.1%	31/57 ms
GPT-4+Memory	0.478±0.013	14.2%	48.3%	1.85/4.23 s
Claude-3+ReAct	0.472±0.014	16.8%	54.2%	2.13/5.85 s
GPT-4+Constraints	0.469±0.012	11.3%	38.6%	1.97/4.61 s
VIGIL (ours)	0.487±0.010	6.4%	18.4%	47/98 ms
VIGIL+HardFilter	0.482±0.010	1.2%	6.1%	49/103 ms
Exact-Reencode	*0.493±0.009*	*5.1%*	*14.2%*	–

A.1.5 Constraint, Calibration, and Budget Robustness

The matrix perturbs one operating condition at a time while preserving the top-50 proposal pool. Each method cell reports rank-one commitment utility, Q-CVR (%), trap-entry % (raw events/75), and p99 latency in milliseconds.

Table 10: Robustness matrix across constraint thresholds, calibrated-score shifts, and shortlist sizes. Default settings recur in each one-factor block to make the local comparison explicit.

Condition	SlackPreserve-H3	VIGIL	VIGIL+HardFilter
NoRedundancy 0.45	0.474, 10.2, 4.8 (4/75), 93	0.483, 7.5, 2.0 (2/75), 99	0.480, 1.4, 1.1 (1/75), 104
NoRedundancy 0.50	0.476, 9.4, 4.1 (3/75), 92	0.487, 6.4, 1.4 (1/75), 98	0.482, 1.2, 0.9 (1/75), 103
NoRedundancy 0.55	0.478, 8.8, 3.7 (3/75), 91	0.487, 6.5, 1.8 (1/75), 97	0.484, 1.0, 0.8 (1/75), 102
EntityCoref 0.82	0.478, 8.7, 3.6 (3/75), 91	0.486, 6.3, 1.6 (1/75), 97	0.484, 1.0, 0.8 (1/75), 102
EntityCoref 0.85	0.476, 9.4, 4.1 (3/75), 92	0.487, 6.4, 1.4 (1/75), 98	0.482, 1.2, 0.9 (1/75), 103
EntityCoref 0.88	0.473, 10.6, 4.9 (4/75), 93	0.480, 7.5, 2.3 (2/75), 99	0.479, 1.6, 1.2 (1/75), 104
TopicCoherence 0.27	0.479, 8.5, 3.5 (3/75), 91	0.489, 5.8, 1.3 (1/75), 97	0.485, 0.9, 0.7 (1/75), 102
TopicCoherence 0.30	0.476, 9.4, 4.1 (3/75), 92	0.487, 6.4, 1.4 (1/75), 98	0.482, 1.2, 0.9 (1/75), 103
TopicCoherence 0.33	0.472, 10.8, 5.1 (4/75), 93	0.476, 11.2, 3.1 (2/75), 99	0.478, 2.1, 1.4 (1/75), 104
Score shift -0.05	0.472, 10.9, 5.0 (4/75), 93	0.483, 7.4, 2.0 (2/75), 99	0.478, 1.7, 1.3 (1/75), 104
Score shift 0	0.476, 9.4, 4.1 (3/75), 92	0.487, 6.4, 1.4 (1/75), 98	0.482, 1.2, 0.9 (1/75), 103
Score shift $+0.05$	0.479, 8.3, 3.5 (3/75), 91	0.490, 5.5, 1.1 (1/75), 97	0.485, 0.9, 0.7 (1/75), 102
$K = 8$	0.462, 13.2, 6.9 (5/75), 74	0.477, 8.6, 2.7 (2/75), 81	0.470, 1.9, 1.4 (1/75), 85
$K = 16$	0.476, 9.4, 4.1 (3/75), 92	0.487, 6.4, 1.4 (1/75), 98	0.482, 1.2, 0.9 (1/75), 103
$K = 32$	0.482, 7.8, 3.0 (2/75), 132	0.491, 5.4, 0.9 (1/75), 144	0.487, 0.9, 0.6 (1/75), 151

VIGIL has lower Q-CVR than SlackPreserve-H3 in 8/9 threshold settings; HardFilter retains the lower Q-CVR operating point in all nine. Across the settings, the median paired differences for VIGIL are -2.7 pp [-4.4,-1.1] Q-CVR and +0.009 [+0.002,+0.015] commitment utility. The advantage grows at $K = 8$, where feasible slack is scarcest, and remains positive at $K = 32$.

A.2 Additional Reliability Evidence

Table 11 shows a smooth operating frontier. Increasing the threshold from 0.65 to 0.85 lowers Q-CVR from 3.4% to 0.8% and p99 from 11.8% to 4.7%, while commitment utility declines gradually from 0.486 to 0.479. The selected 0.80 threshold retains most utility and reaches the low-violation regime used in the full-fit comparison.

Table 11: Hard-filter threshold sweep. The default threshold τ_{hard}=0.8 is the operating point reported in the main paper.

τ_{hard}	Commitment utility	Q-CVR	p99
0.65	0.486±0.010	3.4%	11.8%
0.75	0.484±0.010	2.0%	8.2%
0.80	0.482±0.010	1.2%	6.1%
0.85	0.479±0.011	0.8%	4.7%

Figure 2 uses exactly the 18 official CAsT 2022 information needs in the full-fit analysis. VIGIL records 12/4/2 wins/losses/ties against E5-large, with a mean paired commitment-utility gain of +0.046 [0.031,0.061].

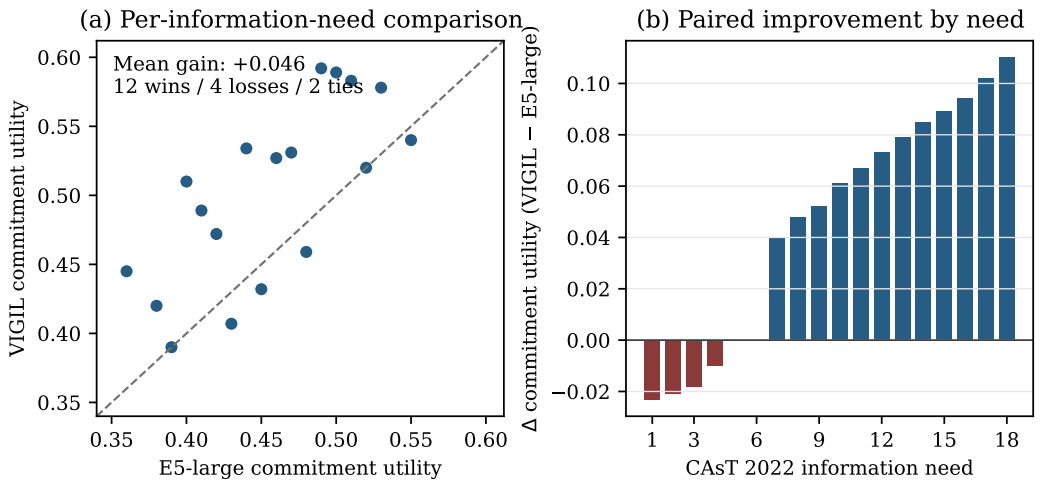


Figure 2: Full-fit commitment utility across all 18 CAsT 2022 information needs. VIGIL records 12/4/2 wins/losses/ties against E5-large and a paired mean gain of +0.046 [0.031,0.061].

VIGIL's gains occur across the information-need distribution rather than being produced by one outlying topic. The paired view also exposes the four negative and two tied topics alongside the twelve positive differences.

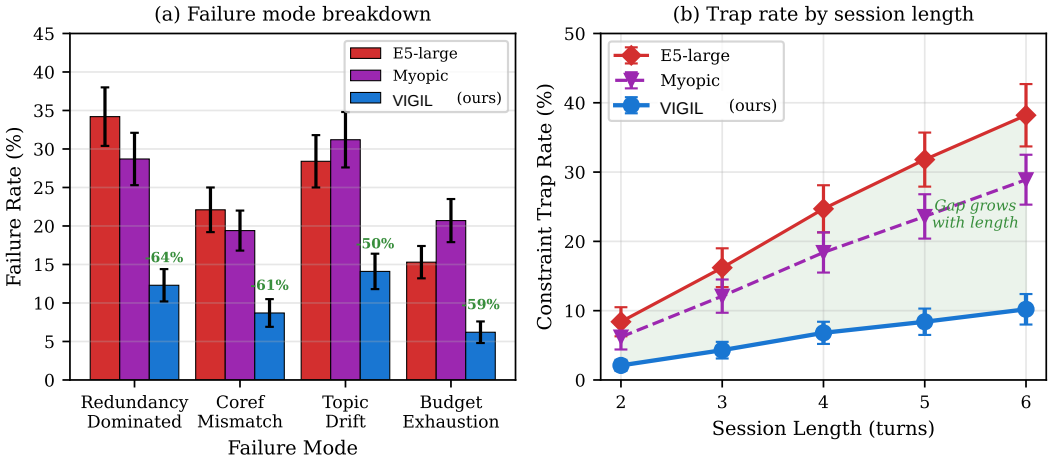


Figure 3: Descriptive full-fit diagnostics over the 75 shared segments. VIGIL reduces all four failure categories, and its observed trap-rate gap over E5-large widens with session length. "Myopic" denotes the $H{=}1$ ablation; "budget exhaustion" means that the fixed $K{=}16$ pool contains no remaining rule-compatible continuation. Bar annotations show relative reductions from E5-large, and whiskers summarize variation across the 18 information needs.

The category reductions and session-length trend connect the aggregate Q-CVR result to two concrete signatures: fewer redundancy, entity, topic, and exhaustion failures, and a larger separation on longer conversational segments.

The prediction diagnostics support the selected three-step score from two directions. Current-action constraint probabilities have aggregate ECE 0.024, while recursive successor-latent fit decreases smoothly from $R^2{=}0.91$ at pseudo-step 1 and remains above $R^2{=}0.8$ at $H{=}3$ before reaching $R^2{=}0.72$ at step 5. The behavioral control is consistent with that choice: reducing the score to $H{=}1$ raises Q-CVR from 6.4% to 15.4%.

A.3 PROOF SKETCHES

Proposition 1 states the current-action property of HardFilter under exact feasibility prediction.

Proposition 1. *At any turn where the thresholded set $\widehat{F}_t = \{a \in \mathcal{A}_t : \min_m \hat{c}_m(z_{t,0}, a) \geq 0.8\}$ is non-empty, if membership in $\widehat{F}_t$ exactly matches current-action feasibility, then VIGIL+HardFilter selects an action in $F(s_t, \mathcal{A}_t)$.*

Proof. HardFilter maximizes Equation 8 over the non-empty set $\widehat{F}_t$. Exact membership gives $\widehat{F}_t = F(s_t, \mathcal{A}_t)$, so its selected action is feasible at the current state. □

For non-fallback turns, define $E_t = \{\widehat{F}_t \neq \varnothing,\ a_t \notin F(s_t, \mathcal{A}_t)\}$, the selected-action false-feasibility event. If $\Pr(E_t \mid s_t) \leq \epsilon_t$, then the union bound gives $\Pr(\bigcup_{t=1}^{T} E_t) \leq \sum_{t=1}^{T} \epsilon_t$ without an independence assumption. Fallback turns, on which $\widehat{F}_t = \varnothing$, are outside this statement and remain visible in the empirical Q-CVR and trap statistics.