ACTS: CONTROLLING EVIDENCE ACQUISITION BEFORE WRITING QUESTIONS FOR SEARCH AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Training search agents for open-world problems requires high-quality long-horizon trajectories that remain scarce and costly. The bottleneck begins before teacher rollout: a synthetic question can appear semantically complex yet collapse to one lookup when its clues reside on a single page, while a larger evidence graph may add redundancy without requiring new retrieval. Existing synthesis methods control end-product properties such as hop count, graph size, or question-level difficulty, but not the acquisition process realized during task construction. This creates a *search-demand gap* between apparent complexity and realized acquisition. We introduce ACTS (*Acquisition-Controlled Task Synthesis*), which designs a task's evidence-acquisition topology before writing its question. From construction-time retrieval, ACTS builds linked entity and document-navigation hypergraphs that record grounded evidence and the queries and documents that made it reachable. Complementary search policies expand a shared state, and a lexicographic max–min objective selects the branch subset that maximizes the weaker of its semantic and acquired-evidence scores, with mean quality and nonredundancy as later tie-breakers. ACTS separately chooses evidence for question writing and the next construction frontier, preserving exploration without exposing every observation to the writer. Generated questions are screened for answer leakage and no-tool shortcuts, then independently solved by a tool-using teacher. In matched construction, ACTS increases mean committed evidence sources from 4.50 to 9.69 while achieving a 70.17% teacher solve rate. Under anti-contamination evaluation, ACTS-trained models reach 67.61% on full BrowseComp and 52.13% on HLE-Text with SFT alone, demonstrating that acquisition-controlled synthesis produces effective supervision across diverse search settings.

1 INTRODUCTION

Search-agent training data must teach a model not only to reason over retrieved facts, but also to discover the right facts through tools. Consider a question that connects three clues. If all three appear on the same page, a teacher may solve the question after one lookup. If each intermediate document instead reveals the entity or query needed to find the next clue, the same semantic chain requires iterative querying, page inspection, and evidence integration. The two questions may look equally multi-hop, but they elicit very different search behavior. Because supervised fine-tuning (SFT) can supervise only the decisions realized in teacher trajectories, the bottleneck begins before teacher rollout: task construction must shape both what evidence supports a question and how that evidence becomes available through search.

Existing pipelines construct search tasks through recursive crawling, easy-to-hard query evolution, entity walks, graph expansion, and formal query construction (Wu et al., 2025; Li et al., 2025b; Lu et al., 2025; Xie et al., 2026; Tao et al., 2025; Liu et al., 2025). Recent methods also control logical topology and source dispersion, screen shortcut risk, or verify task and trajectory evidence (Chu et al., 2026; Deng et al., 2026; Mei et al., 2026). These controls shape the evidence scaffold from which a question is written, but they do not fully determine the evidence-acquisition process that the task will elicit. Even when retrieval is used during construction, its outcomes are not jointly selected with semantic structure as an explicit synthesis objective. The same final evidence graph may therefore arise from a shallow lookup, repeated retrieval, or a sequence in which one document

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.

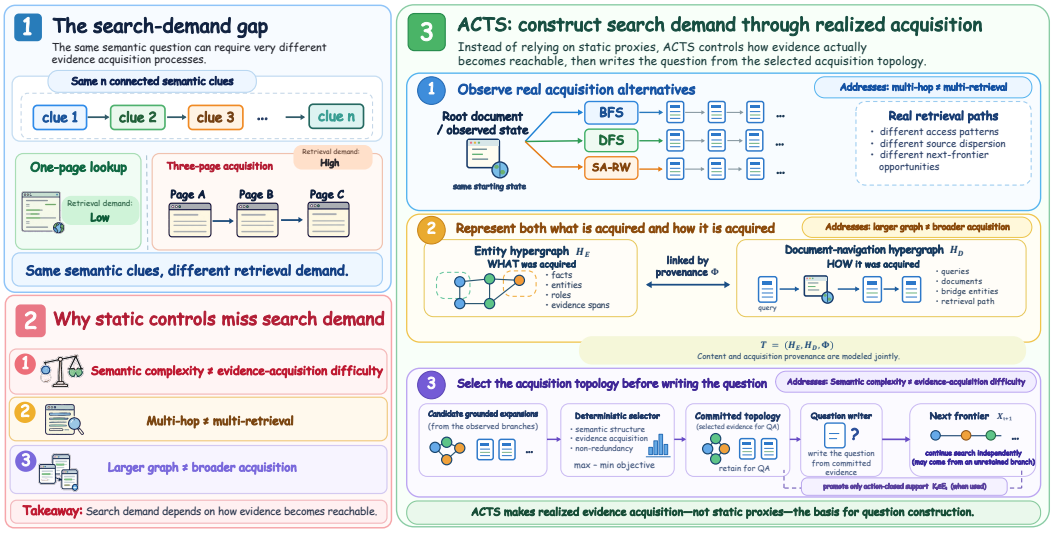


Figure 1: Static task proxies leave search demand under-specified. Questions with similar semantic structure can require different retrieval processes, while additional hops or graph elements need not broaden evidence acquisition. ACTS therefore constructs and selects a grounded evidence-acquisition topology before question writing.

reveals the bridge to another. We call this mismatch between apparent semantic complexity and realized evidence acquisition the *search-demand gap*.

Figure 1 illustrates why static complexity proxies do not determine realized search demand and how ACTS moves acquisition control before question writing.

We introduce ACTS (*Acquisition-Controlled Task Synthesis*), which turns evidence acquisition from a by-product of construction into an explicit design object before question writing. This requires controlling three distinct decisions: how acquired evidence is represented, which realized branches are retained for writing, and where acquisition continues. A pairwise entity graph flattens an n-ary fact into independent binary edges and omits the query–document actions that exposed it. Its structure therefore cannot distinguish joint relational evidence from its projection or reveal how that evidence became reachable. **Linked Hypergraph Representation (LHR)** addresses both losses: an entity hypergraph preserves grounded facts, participant roles, and evidence spans, while a document-navigation hypergraph records the executed queries, source and retrieved documents, and bridge entities. Provenance links the two views into the evidence-acquisition topology that ACTS controls.

Representation alone does not determine which realized acquisitions should shape the task. A fixed search policy inherits one traversal bias, while committing every explored branch can replace useful breadth with duplicated facts and sources. **Multi-Policy Fusion (MPF)** therefore explores with complementary policies from a shared state and makes the retained policy subset a construction decision. Its lexicographic max–min objective first strengthens the weaker of semantic composition and acquired evidence, then favors quality and nonredundancy. The selector thus retains multiple branches when they add complementary structure, without equating more branches with a better task.

Retention and continuation serve different horizons: retention determines the evidence exposed to the writer now, whereas continuation chooses the frontier that can unlock evidence in the next round. Coupling them either abandons a promising frontier when its branch is not retained or retains a redundant branch merely to keep that frontier alive. **Retention–Continuation Decoupling (RCD)** separates these decisions. Construction can continue from an unretained branch while adding only the support needed to keep the committed evidence grounded and connected.

The writer generates a question from the committed entity hypergraph. Answer-leakage checks and a no-tool solver remove direct-answering cases; an independent tool-using trajectory teacher then

receives only the surviving question and tools and produces the training trajectory through its own search. Figure 2 summarizes this path from controlled evidence acquisition to SFT supervision.

Our experiments follow the same construction-to-learning chain. Relative to a single-policy ordinary-graph baseline, full ACTS increases the mean number of committed evidence sources in judged-correct teacher trajectories from 4.50 to 9.69 and mean teacher tool calls from 27.47 to 31.60, with a 70.17% teacher solve rate. The trajectory teacher, which never observes the construction-time hypergraphs, recovers 59.95% of its non-answer construction-path nodes, compared with less than 1% under mismatched controls. The resulting supervision transfers through SFT across three student backbones and four online-search benchmarks. Under anti-contamination evaluation, ACTS-trained Qwen3.8-27B reaches 67.61% on BrowseComp and 52.13% on HLE-Text, while Qwen3.5-35B-A3B reaches 61.45% on BrowseComp (Wei et al., 2025).

Our contributions are:

- We identify the search-demand gap and formulate search-task synthesis as controlling realized evidence acquisition before question writing. Linked entity and document-navigation hypergraphs jointly represent grounded semantic structure and its acquisition provenance.
- We introduce a parameterized multi-policy constructor with lexicographic max–min selection and retention–continuation decoupling. The formulation recovers single-policy ordinary-graph and hypergraph construction, as well as all-branch union, as special cases.
- We establish the construction-to-learning chain empirically: ACTS broadens committed source evidence while preserving teacher solvability, its construction paths reappear in independent teacher search, and the resulting supervision improves three student backbones across four search benchmarks.

2 RELATED WORK

Synthetic tasks for information-seeking agents. Existing pipelines create training tasks by crawling pages, evolving queries, sampling structured paths, or projecting information needs into formal constraints. WebDancer recursively follows links and composes questions from collected pages; WebSailor expands entity graphs; DeepDive samples knowledge-graph walks; WebExplorer evolves tasks through exploration; and WebShaper formalizes information-seeking objectives (Wu et al., 2025; Li et al., 2025b; Lu et al., 2025; Liu et al., 2025; Tao et al., 2025). OpenSeeker distills entity subgraphs from expanded page graphs (Du et al., 2026b). OpenSeeker-v2 emphasizes informative long trajectories (Du et al., 2026a), while QUEST constructs synthetic tasks with rubric trees (Xie et al., 2026). These methods show that the structure of synthetic tasks matters. ACTS advances this direction by selecting an evidence-acquisition topology before question writing, so the construction record specifies both the evidence and the acquisition actions that made it available.

Structural control of search demand. Multi-hop QA connects facts across passages (Yang et al., 2018; Trivedi et al., 2022; Ho et al., 2020; Talmor & Berant, 2018), while newer synthesis methods explicitly target search difficulty. REDSearcher controls logical topology and source dispersion, FORT-Searcher controls shortcut risks across task construction and diagnoses their effects through trajectory signatures, and SearchArt jointly verifies question–answer consistency, trajectory quality, and retrieved evidence (Chu et al., 2026; Deng et al., 2026; Mei et al., 2026). Their objectives make task structure, shortcut resistance, and verification explicit. ACTS makes the acquisition process itself a construction variable: it compares branch increments produced by different search policies from the same observed state, then selects among them using both semantic content and acquisition provenance.

Evidence graphs and acquisition provenance. Graph-based systems increasingly record how evidence supports reasoning. EviGraph builds typed claim–evidence memory during inference, SearchAtlas reconstructs evidential query graphs from completed trajectories, and Bridge Evidence distinguishes static document relevance from the causal value of a document that enables later retrieval (Chen et al., 2026; Sang et al., 2026; Mukhopadhyay et al., 2026). ACTS moves this representation to the point of task creation. Linked entity and document-navigation hypergraphs jointly represent grounded facts and the acquisition actions that exposed them. This construction-time view

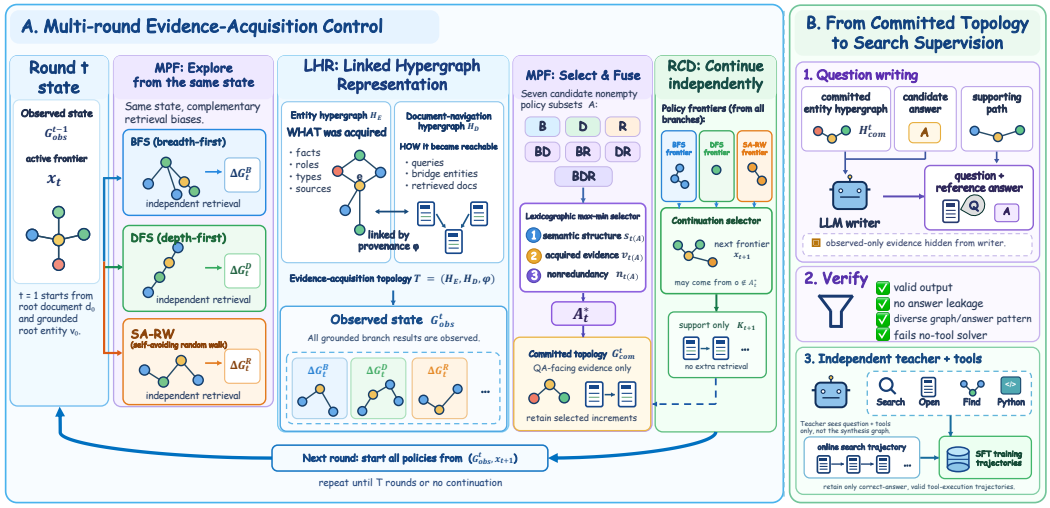


Figure 2: The ACTS construction and supervision pipeline. (A) LHR records linked entity and document-navigation hypergraphs from complementary policy expansions; MPF selects and fuses committed evidence, while RCD chooses the continuation frontier independently. (B) The writer receives committed entity evidence, and the trajectory teacher receives only the resulting question and registered tools.

allows alternative construction paths to be compared before a question is written, rather than reconstructed after a trajectory is complete.

3 ACTS: ACQUISITION-CONTROLLED TASK SYNTHESIS

Figure 2 presents the ACTS pipeline. Given a root document d_0 and a root entity v_0 grounded in it, ACTS extracts grounded facts and related entities, queries the target corpus for documents about those entities, and repeats this process with newly discovered entities. Three core modules control this process: Linked Hypergraph Representation (LHR), Multi-Policy Fusion (MPF), and Retention–Continuation Decoupling (RCD). LHR records the acquired facts, documents, and acquisition actions. MPF compares the evidence acquired by multiple search policies, selects a policy subset with a max–min objective, and fuses its evidence into the committed state. RCD independently chooses the continuation entity for the next round.

After the final round, the writer generates a question and reference answer from the evidence accumulated in the committed entity hypergraph. The generated question is checked for answer leakage and no-tool answerability. A trajectory teacher solves each question that passes these checks with search tools, and the retained teacher trajectories are used to train the student agent through SFT.

3.1 LINKED HYPERGRAPH REPRESENTATION (LHR)

LHR represents the information acquired during search with two linked hypergraphs. The ordinary-graph representation uses a pairwise entity projection that splits a fact such as "A and B co-founded C" into separate A–C and B–C edges and does not represent the acquisition action that exposed the fact. LHR instead stores grounded facts in an entity hypergraph and acquisition actions in a document-navigation hypergraph.

Entity hypergraph. For every visited document, the extractor represents each grounded fact as

$$e = (p, \{(v_i, \rho_i, \tau_i)\}_{i=1}^{a}), \tag{1}$$

where p is the predicate and (v_i, ρ_i, τ_i) gives the identity, role, and type of participant i. The entity hypergraph $\mathcal{H}_{\mathrm{E}} = (V_{\mathrm{E}}, E_{\mathrm{E}})$ stores entities as nodes and facts as hyperedges. Unlike a binary edge ($a = 2$), a hyperedge keeps all participants and their roles in one fact.

Document-navigation hypergraph. When ACTS expands an entity v found in source document d_s, it issues query q_v and retrieves documents $D_v = \{d_1, \ldots, d_m\}$. We record this acquisition action as

$$g = (d_s, q_v, v, D_v). \tag{2}$$

The document-navigation hypergraph $\mathcal{H}_{\mathrm{D}} = (V_{\mathrm{D}}, E_{\mathrm{D}})$ stores documents as nodes and each action g as a hyperedge from d_s to the retrieved documents in D_v. ACTS extracts facts only from a result whose normalized title matches v, preventing evidence from unrelated or namesake entities from entering $\mathcal{H}_{\mathrm{E}}$. Source-document identifiers on facts and bridge-entity identifiers on actions link the two hypergraphs. Together with these provenance links, the linked hypergraphs form an *evidence-acquisition topology*: the entity view records what was acquired, while the document-navigation view records how it became reachable. Appendix D provides an example. This separation prevents relational complexity from being conflated with acquisition breadth. Under a fixed branch trace, preserving native n-ary facts should retain higher-order structure at the same document count.

A parameterized constructor family. We write an acquisition-controlled constructor as

$$\mathrm{ACTS}(\varrho, \mathcal{O}, \mathfrak{A}, \mathbf{b}, \mathsf{Cont}), \tag{3}$$

where ϱ specifies the evidence representation, $\mathcal{O}$ the search policies, $\mathfrak{A} \subseteq 2^{\mathcal{O}} \setminus \{\emptyset\}$ the admissible policy subsets, $\mathbf{b}$ their policy expansion budgets, and Cont the continuation rule. Restricting $\mathcal{O}$ to SA-RW gives ordinary-graph or hypergraph SA-RW according to ϱ; setting $\mathfrak{A} = \{\mathcal{O}\}$ commits all policy branches. Full ACTS uses the linked hypergraphs, all nonempty policy subsets of BFS, DFS, and SA-RW, and a separate continuation rule.

3.2 MULTI-POLICY FUSION (MPF)

Each construction starts from a root $r = (d_0, v_0)$, where d_0 is a document in the KILT Wikipedia corpus (Petroni et al., 2021) and v_0 is an entity grounded in d_0. Before the first of at most T construction rounds, ACTS extracts facts from d_0. These facts initialize the observed state $\mathcal{G}_0^{\mathrm{obs}} = (\mathcal{H}_{\mathrm{E},0}^{\mathrm{obs}}, \mathcal{H}_{\mathrm{D},0}^{\mathrm{obs}})$, and their non-root entities provide the first search candidates.

Multi-policy exploration. At round t, BFS, DFS, and SA-RW each perform at most H policy expansion steps, producing three branch traces. BFS favors broader coverage, DFS follows deeper paths, and SA-RW explores unvisited entities stochastically. The root entity is active in the first round, and candidates extracted from d_0 initialize the three policy frontiers; thereafter, the preceding round's continuation entity becomes the shared active entity. The policies share the same input but execute independently within the round.

For each policy $o \in \mathcal{O}$, let

$$\Delta\mathcal{G}_t^o = (\Delta\mathcal{H}_{\mathrm{E},t}^o, \Delta\mathcal{H}_{\mathrm{D},t}^o) \tag{4}$$

denote the content acquired for the entity and document-navigation hypergraphs in round t. After the three branches finish, ACTS adds all three branch increments to the observed state:

$$\mathcal{G}_t^{\mathrm{obs}} = \mathcal{G}_{t-1}^{\mathrm{obs}} \cup \bigcup_{o \in \mathcal{O}} \Delta\mathcal{G}_t^o. \tag{5}$$

The observed state retains all acquired information for subsequent search, but not all of it is used to generate the question.

Max–min selection. A fixed policy can miss evidence exposed by another policy, whereas taking the union of all three branches can introduce redundant facts and documents. MPF therefore considers seven candidate policy subsets: each singleton, each pair, and the set of all three policies. Let $A \subseteq \mathcal{O}$ denote one such subset. For each branch, ACTS derives four feature sets from its branch trace: relation patterns $Z_{\mathrm{sch},t}^o$, grounded entity–role instances $Z_{\mathrm{con},t}^o$, evidence documents $Z_{\mathrm{doc},t}^o$, and grounded facts $Z_{\mathrm{fact},t}^o$. Their exact construction is given in Appendix A.

Let $Z_{k,t}(A) = \bigcup_{o \in A} Z_{k,t}^o$. For each channel $k \in \{\mathrm{sch}, \mathrm{con}, \mathrm{doc}, \mathrm{fact}\}$, ACTS balances coverage of the full round with efficiency within the selected policy branches:

$$r_{k,t}(A) = \frac{1}{2}\left(\frac{|Z_{k,t}(A)|}{|Z_{k,t}(\mathcal{O})|} + \frac{|Z_{k,t}(A)|}{\sum_{o \in A} |Z_{k,t}^o|}\right), \tag{6}$$

where a ratio with a zero denominator is zero. The first term rewards retained information; the second decreases with cross-branch duplication.

Let ℓ_t^o be the realized depth of branch o and $h_t(A) = \min\{\max_{o\in A} \ell_t^o/H, 1\}$. The semantic, evidence, and nonredundancy scores are

$$s_t(A) = \tfrac{1}{3}\left(h_t(A) + r_{\mathrm{sch},t}(A) + r_{\mathrm{con},t}(A)\right), \tag{7}$$

$$v_t(A) = \tfrac{1}{2}\left(r_{\mathrm{doc},t}(A) + r_{\mathrm{fact},t}(A)\right), \tag{8}$$

$$n_t(A) = \tfrac{1}{4} \sum_{k\in\{\mathrm{sch,con,doc,fact}\}} \frac{|Z_{k,t}(A)|}{\sum_{o\in A} |Z_{k,t}^o|}. \tag{9}$$

MPF selects the retained policy subset with

$$A_t^* = \operatorname*{lexmax}_{\emptyset\neq A\subseteq\mathcal{O}} \left(\min\{s_t(A), v_t(A)\}, \frac{s_t(A)+v_t(A)}{2}, n_t(A), -|A| \right). \tag{10}$$

The objective first balances semantic structure and acquired evidence, then favors higher mean quality, lower duplication, and fewer branches. Its properties are analyzed in Appendix A. Accordingly, the selector should favor multi-policy subsets when branches contribute complementary evidence and smaller subsets as their evidence overlaps.

Given A_t^*, MPF's fusion operator returns $\Phi_t = \bigcup_{o\in A_t^*} \Delta\mathcal{G}_t^o$. It updates the committed state, initialized by $\mathcal{G}_0^{\mathrm{com}} = \mathcal{G}_0^{\mathrm{obs}}$:

$$\mathcal{G}_t^{\mathrm{com}} = \mathcal{G}_{t-1}^{\mathrm{com}} \cup \mathcal{K}_t \cup \Phi_t, \qquad \mathcal{K}_1 = \emptyset, \tag{11}$$

where $\mathcal{K}_t$ contains all recorded support for the active entity from which round t starts. Only content from the committed entity hypergraph is exposed to the writer.

3.3 Retention–Continuation Decoupling (RCD)

The evidence best suited to question writing need not contain the most promising frontier for continued acquisition. RCD therefore chooses continuation separately from retention. Each policy exposes a frontier of unexpanded entities. Among policies with nonempty frontiers, RCD first prefers a continuation entity outside the committed entity hypergraph. Remaining ties are resolved by frontier novelty and width, realized depth, and the singleton bottleneck score. We write this deterministic choice as

$$o_t^* = \operatorname*{lexmax}_{o\in\mathcal{O}} \mathbf{c}_t^o, \qquad x_{t+1} = \mathrm{next}(o_t^*), \tag{12}$$

where the exact priority vector $\mathbf{c}_t^o$ is defined in Appendix A. Search stops if every policy frontier is empty.

The selected continuation entity can come from a branch outside A_t^*. Rather than committing that branch in full, ACTS adds only the support already recorded along the path to x_{t+1}:

$$\mathcal{K}_{t+1} = \mathrm{Supp}_{\mathrm{obs}}(x_{t+1}) \setminus \mathcal{G}_t^{\mathrm{com}}. \tag{13}$$

This requires no additional retrieval and keeps the committed state connected to the root (Proposition 2). The next round starts all policies from the same $\mathcal{G}_t^{\mathrm{obs}}$ and x_{t+1}. In this way, construction can explore beyond the retained policy subset while keeping the writer's evidence grounded and connected.

ACTS stops after T rounds or when no continuation entity is available. The full configuration uses $H = 3$ and $T = 3$, for at most $3HT = 27$ policy expansion steps. Appendix Figure 3 summarizes the complete procedure.

3.4 Question writing and teacher rollout

After the final round, ACTS selects a grounded non-root entity as the candidate answer. The writer receives only the committed entity hypergraph, the candidate answer, and its construction path. It produces a question and reference answer.

Deterministic checks remove malformed outputs and questions that reveal the answer. A diversity filter limits repeated graph and answer patterns, and a solver without tools removes questions

answerable from parametric knowledge alone. The trajectory teacher then receives the surviving question and registered tools, but not the construction-time hypergraphs. Judged-correct teacher trajectories with valid tool execution are retained for supervised fine-tuning. Figure 2 summarizes this information flow; Appendix B.1 specifies model roles and information access.

4 EXPERIMENTS

4.1 BENCHMARKS AND EVALUATION PROTOCOL

We evaluate the complete 1,266-question BrowseComp (BC) test set (Wei et al., 2025), the 100-question xBench-DeepSearch-2510 (xBench-2510) set (Chen et al., 2025), the 111-question SEAL-0 set (Pham et al., 2026), and the 2,158-question HLE-Text set (Phan et al., 2026). Each evaluation provides online search, page opening, in-page finding, and sandboxed computation. We actively monitor for potential benchmark contamination and block access to identified sources at the infrastructure level; all ReAct and ACTS results use this protocol. GPT-5.5 judges answer equivalence over each complete benchmark denominator. Appendix C.1 gives the evaluation protocols. The matched ACTS–QUEST comparison uses the 130-question BrowseComp subset from QUEST and a larger 200-question diagnostic set; the six-arm construction ablation uses the same fixed 130-question split across all arms.

4.2 END-TO-END TRANSFER ACROSS STUDENTS AND BENCHMARKS

We train Qwen3-30B-A3B-Thinking-2507 (Qwen3-30B), Qwen3.8-27B, and Qwen3.5-35B-A3B (Qwen3.5-35B) on the same 11,651 curated Kimi-K3 trajectories generated from ACTS questions (564,364 structured tool calls; 48.44 per trajectory). We additionally report a 16,792-trajectory multi-teacher Qwen3-30B recipe marked with $\star$. Appendix B documents the recipes, multi-teacher composition, and curation criteria.

Table 1 compares representative full-benchmark systems, including SFT-only agents, broader post-training pipelines, and proprietary reference systems.

Under anti-contamination evaluation, the same 11,651-trajectory training set improves every matched ReAct baseline on all four benchmarks. The gains are not merely inherited from strong backbones: on BrowseComp, Qwen3 rises from 2.61% to 43.76%, while Qwen3.5 and Qwen3.8 gain 20.53 and 18.56 points over their respective bases. Qwen3.8 translates this supervision into the strongest performance among the ACTS students, remaining competitive with systems that use broader post-training while relying on SFT alone. For Qwen3, the broader 16,792-trajectory multi-teacher recipe improves all four benchmarks over the 11,651-trajectory recipe. These complementary results show both substantial learning gains from weaker starting points and strong transfer as student capability increases.

4.3 HOW CONSTRUCTION CHOICES SHAPE ACQUISITION AND LEARNING

The six-arm construction ablation connects construction choices to trajectory behavior and downstream SFT. All arms share 640 root pairs $r = (d_0, v_0)$, the KILT snapshot (Petroni et al., 2021), $H = 3, T = 3$, and the same generation settings. We analyze 1,800 complete trajectories per arm and train each student on 2,000 judged-correct trajectories under the same SFT protocol and BC-130 split. The ordinary-graph/hypergraph pair isolates representation; the multi-policy arms compare retention and fusion rules; and the size-matched random control matches the full selector's subset cardinality.

Multi-policy construction increases mean evidence sources from 4.50–4.63 to 9.34–10.76, with teacher solve rates of 69–70%. Holding the SA-RW branch trace fixed, replacing the ordinary-graph representation with the entity hypergraph changes the fraction with treewidth ≥ 4 from 0.00% to 31.34%, while other metrics remain close. Among the multi-policy variants, full ACTS has the highest teacher solve rate and call count. Compared with the entity-only selector, incorporating document evidence yields 9.69 versus 9.34 sources and 57.01% versus 54.47% with treewidth ≥ 4. On BC-130, hypergraph SA-RW improves over ordinary-graph SA-RW, every multi-policy variant improves over both single-policy controls, and full ACTS performs best at 45.38%. All-branch union

Table 1: Full-benchmark online search-agent accuracy (%). Published rows use their reported inference policies; ReAct and ACTS use our anti-contamination protocol. Slashes: without/with context management; MT: mid-training; HLE: HLE-Text; –: unreported; ⋆: multi-teacher Qwen3 (Appendix B.3).

Method	Base	Training	SFT data	BC	xBench-2510	SEAL-0	HLE
Proprietary agents and models							
OpenAI Deep Research	Proprietary	–	–	51.5	–	–	26.6
GPT-5	Proprietary	–	–	54.9	75.0	51.4	35.2
Gemini 3 Pro	Proprietary	–	–	59.2	53.0	45.5	46.9
Claude Opus 4.5	Proprietary	–	–	67.8	–	47.7	43.2
Claude 4 Sonnet	Proprietary	–	–	12.2	35.0	–	–
Open models and search agents							
DeepDive-32B (Lu et al., 2025)	QwQ-32B	SFT + RL	858	15.3	–	–	–
LightThinker++ vanilla (Zhu et al., 2026)	Qwen3-30B	SFT	6,625	16.0	38.3	–	–
LightThinker++ (Zhu et al., 2026)	Qwen3-30B	SFT	3,677	18.1	44.0	–	–
WebSailor-V2 (Li et al., 2025a)	Qwen3-30B	SFT + RL	–	35.3	–	–	30.6
WebResearcher (Qiao et al., 2025)	Qwen3-30B	SFT + RL	–	37.3/51.7	–	–	28.8/36.7
OpenResearcher (Li et al., 2026)	Nemotron-30B	SFT	55K	26.3	–	–	19.6
QUEST-30B (Xie et al., 2026)	Qwen3-30B	MT + SFT + RL	23.9K	37.0	–	–	24.6
Tongyi-DeepResearch (Tongyi DeepResearch Team et al., 2026)	Qwen3-30B	MT + SFT + RL	–	43.4	55.0	45.8	32.9
REDSearcher (Chu et al., 2026)	Qwen3-30B	MT + SFT + RL	10K	42.1/57.4	–	–	34.3
FORT-Searcher (Deng et al., 2026)	Qwen3-30B	SFT	–	55.9/72.2	54.1/57.2	43.7/46.0	–
MiroThinker-1.7-mini (MiroMind Team et al., 2026)	Qwen3-30B	MT + SFT + DPO + RL	–	67.9	57.2	48.2	36.4
SearchArt (Mei et al., 2026)	Qwen3.5-27B	SFT + RL	–	70.06	–	–	–
QUEST-35B (Xie et al., 2026)	Qwen3.5-35B	MT + SFT + RL	23.9K	45.5/64.6	–	–	37.2
ReAct baselines under anti-contamination evaluation							
ReAct (Yao et al., 2023)	Qwen3-30B	–	–	2.61	5.00	7.21	9.27
ReAct (Yao et al., 2023)	Qwen3.5-35B	–	–	40.92	49.00	39.64	30.72
ReAct (Yao et al., 2023)	Qwen3.8-27B	–	–	49.05	61.00	50.45	47.78
ACTS under anti-contamination evaluation (SFT only)							
ACTS	Qwen3-30B	SFT	11,651	43.76	39.00	31.53	26.83
ACTS	Qwen3-30B	SFT	16,792⋆	50.47	45.00	33.33	32.48
ACTS	Qwen3.5-35B	SFT	11,651	61.45	56.00	48.65	47.36
ACTS	**Qwen3.8-27B**	**SFT**	**11,651**	**67.61**	**67.00**	**58.56**	**52.13**

Table 2: Six-arm construction ablation; metric definitions and denominators appear in Appendix C.2.

	Construction design			Trajectory diagnostics					Downstream
Arm	Entity repr.	Policies	Retention/ fusion	Tool-required	Teacher solve	Calls	Sources	TW ≥ 4	BC-130
Ordinary-graph SA-RW	Pairwise	SA-RW	Single	32.12	**77.06**	27.47	4.50	0.00	36.15
Hypergraph SA-RW	Hypergraph	SA-RW	Single	31.80	76.22	28.71	4.63	31.34	37.69
All-branch union	Hypergraph	BFS/DFS/SA-RW	Commit all	**34.59**	69.39	31.10	**10.76**	**58.69**	41.54
Entity-only selector	Hypergraph	BFS/DFS/SA-RW	Entity max–min	34.51	69.56	30.69	9.34	54.47	43.08
Size-matched random	Hypergraph	BFS/DFS/SA-RW	Random-k	33.77	69.33	30.04	9.82	52.24	42.31
ACTS (full)	Hypergraph	BFS/DFS/SA-RW	Entity+doc. max–min	33.28	70.17	**31.60**	9.69	57.01	**45.38**

yields the broadest raw source set, yet full ACTS gains 3.84 BC-130 points while committing 1.07 fewer sources, showing that selective complementarity matters more than unfiltered breadth.

4.4 MATCHED COMPARISON WITH QUEST

To isolate the data recipe, we compare similarly sized Kimi-K3 trajectory sets from ACTS and QUEST (Xie et al., 2026) under the same tools, Qwen3.8-27B student, six-epoch SFT schedule, and evaluation protocol. Appendix Table 11 gives the complete configuration.

At comparable training-trajectory counts, the ACTS-trained student exceeds the QUEST-trained student by 8.46 percentage points on QUEST's BC-130 split. Both students score lower on the larger BC-200 diagnostic, where the gap grows to 13.5 points. At similar row counts, ACTS supplies 1.79×

as many tool calls; its larger BC-200 advantage shows that richer acquisition supervision transfers beyond QUEST's original split.

Table 3: Matched ACTS–QUEST comparison on QUEST BC-130 and BC-200.

Training recipe	SFT trajectories	Tool calls	Epochs	QUEST BC-130	BC-200 diagnostic
ACTS (ours)	4,670	243,613	6	89/130 (68.46%)	130/200 (65.00%)
QUEST	4,516	135,898	6	78/130 (60.00%)	103/200 (51.50%)

4.5 FUSION ADAPTS TO BRANCH REDUNDANCY ACROSS ROUNDS

To test whether MPF adapts as branches converge, we track its retained subsets and cross-branch overlap across 17,734 construction decisions with complete scores for all seven candidate policy subsets.

Table 4: SFT-weighted MPF statistics by round (%; $N = 6{,}776/6{,}623/4{,}335$).

Selector statistic	Round 1	Round 2	Round 3	Overall
BFS	14.30	51.50	47.43	36.29
DFS	9.58	16.25	19.10	14.40
SA-RW	5.11	12.70	14.09	10.14
BFS+DFS	10.80	0.27	0.09	4.25
BFS+SA-RW	14.79	8.61	8.26	10.88
DFS+SA-RW	32.97	9.72	9.90	18.65
BFS+DFS+SA-RW	12.46	0.95	1.13	5.39
Any multi-policy subset	71.02	19.55	19.38	39.17
Mean cross-branch overlap	40.00	57.49	56.94	50.67

Mean cross-branch overlap across the selector's four feature channels rises from 40.00% in round 1 to 57.49% in round 2, while singleton retention rises from 28.98% to 80.45%. Among constructions reaching both adjacent rounds, overlap increases by 17.92 points from round 1 to 2 and by 4.18 points from round 2 to 3. Correspondingly, multi-policy retention falls from 71.02% to 19.55% and 19.38%, consistent with the selector's coverage and nonredundancy terms. Evidence retention and continuation selection serve different purposes: RCD selects a continuation policy outside the retained policy subset in 4,027/10,958 transitions (36.75%). Definitions, channel-level results, and unique-construction sensitivity appear in Appendix B.4.

4.6 CONSTRUCTION PATHS REAPPEAR IN TEACHER SEARCH

The central link is whether a construction path remains reflected in the question and in the search behavior of a solver that never sees the construction-time hypergraphs. We link all 11,651 curated trajectories to their source constructions, remove the answer node and its source documents, and compare each pair with two mismatched-construction controls that preserve the question and trajectory while replacing the associated construction. The final questions represent 62.78% of their non-answer path nodes, and 82.52% represent at least two. From the question alone, the Kimi-K3 trajectory teacher searches 51.04% of the construction bridge terms and recovers 59.95% of the non-answer path nodes; all corresponding controls are below 1%. Evidence from the construction paths is therefore reflected in both the written tasks and teacher search. Appendix B.5 gives the full metrics, controls, and matching rules.

5 CONCLUSION

Semantic complexity does not determine what a search agent must retrieve. ACTS addresses this search-demand gap by constructing an evidence-acquisition topology before question writing. It records both grounded evidence and how that evidence became reachable, compares alternative construction paths before deciding what enters the question, and continues construction from promising frontiers beyond the committed evidence. The six-arm construction ablation shows that these choices preserve higher-order structure, broaden evidence-source coverage, and adapt retention to cross-branch redundancy. Large-scale analyses further show that the construction paths reappear in both final questions and teacher trajectories. With SFT alone, the resulting supervision transfers across three student backbones and four online-search benchmarks; under anti-contamination evaluation, Qwen3.8-27B reaches 67.61% on BrowseComp. These results establish evidence acquisition as a controllable and useful object for search-task synthesis.

AI USE STATEMENT

We used generative AI tools only in a limited, ancillary manner for minor language polishing and preliminary figure drafts. All references were independently reviewed and verified by the authors. The authors revised and finalized all AI-assisted material and take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

The appendix specifies the method definitions, model roles, data provenance, curation thresholds, training recipes, benchmark protocols, and metric denominators. Per-example provenance records connect root pairs, acquisition actions, observed and committed states, writer inputs, filter decisions, teacher trajectories, judgments, and SFT membership. We will release the code, configurations, and auditable data artifacts needed to reproduce the reported analyses.

REFERENCES

Jiashun Chen, Yirong Mao, and Wenhui Que. Evigraph: Towards verifiable evidence construction for information-seeking agents, 2026. URL https://arxiv.org/abs/2608.24667.

Kaiyuan Chen, Yixin Ren, Yang Liu, Xiaobo Hu, Haotong Tian, Tianbao Xie, Fangfu Liu, Haoye Zhang, Hongzhang Liu, Yuan Gong, Chen Sun, Han Hou, Hui Yang, James Pan, Jianan Lou, Jiayi Mao, Jizheng Liu, Jinpeng Li, Kangyi Liu, Kenkun Liu, Rui Wang, Run Li, Tong Niu, Wenlong Zhang, Wenqi Yan, Xuanzheng Wang, Yuchen Zhang, Yi-Hsin Hung, Yuan Jiang, Zexuan Liu, Zihan Yin, Zijian Ma, and Zhiwen Mo. xbench: Tracking agents productivity scaling with profession-aligned real-world evaluations, 2025. URL https://arxiv.org/abs/2506.13651.

Zheng Chu, Xiao Wang, Jack Hong, Huiming Fan, Yuqi Huang, Yue Yang, Guohai Xu, Chenxiao Zhao, Cheng Xiang, Shengchao Hu, Dongdong Kuang, Ming Liu, Bing Qin, and Xing Yu. Redsearcher: A scalable and cost-efficient framework for long-horizon search agents, 2026. URL https://arxiv.org/abs/2602.14234.

Jia Deng, Yimeng Chen, Xiaoqing Xiang, Ziyang Zeng, Shuo Tang, Wayne Xin Zhao, Feng Chang, Chuan Hao, Yuan Wei, Ran Tao, Bryan Dai, and Ji-Rong Wen. Fort-searcher: Synthesizing shortcut-resistant search tasks for training deep search agents, 2026. URL https://arxiv.org/abs/2606.12087.

Yuwen Du, Rui Ye, Shuo Tang, Keduan Huang, Xinyu Zhu, Yuzhu Cai, and Siheng Chen. Openseeker-v2: Pushing the limits of search agents with informative and high-difficulty trajectories, 2026a. URL https://arxiv.org/abs/2605.04036.

Yuwen Du, Rui Ye, Shuo Tang, Xinyu Zhu, Yijun Lu, Yuzhu Cai, and Siheng Chen. Openseeker: Democratizing frontier search agents by fully open-sourcing training data, 2026b. URL https://arxiv.org/abs/2603.15594.

Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara, and Akiko Aizawa. Constructing a multi-hop QA dataset for comprehensive evaluation of reasoning steps, 2020. URL https://arxiv.org/abs/2011.01060.

Kuan Li, Zhongwang Zhang, Huifeng Yin, Rui Ye, Yida Zhao, Liwen Zhang, Litu Ou, Dingchu Zhang, Xixi Wu, Jialong Wu, Xinyu Wang, Zile Qiao, Zhen Zhang, Yong Jiang, Pengjun Xie, Fei Huang, and Jingren Zhou. Websailor-v2: Bridging the chasm to proprietary agents via synthetic data and scalable reinforcement learning, 2025a. URL https://arxiv.org/abs/2509.13305.

Kuan Li, Zhongwang Zhang, Huifeng Yin, Liwen Zhang, Litu Ou, Jialong Wu, Wenbiao Yin, Baixuan Li, Zhengwei Tao, Xinyu Wang, Weizhou Shen, Junkai Zhang, Dingchu Zhang, Xixi Wu, Yong Jiang, Ming Yan, Pengjun Xie, Fei Huang, and Jingren Zhou. Websailor: Navigating superhuman reasoning for web agent, 2025b. URL https://arxiv.org/abs/2507.02592.

Zhuofeng Li, Dongfu Jiang, Xueguang Ma, Haoxiang Zhang, Ping Nie, Yuyu Zhang, Kai Zou, Jianwen Xie, Yu Zhang, and Wenhu Chen. OpenResearcher: A fully open pipeline for long-horizon deep research trajectory synthesis, 2026. URL https://arxiv.org/abs/2603.20278v1.

Junteng Liu, Yunji Li, Chi Zhang, Jingyang Li, Aili Chen, Ke Ji, Weiyu Cheng, Zijia Wu, Chengyu Du, Qidi Xu, Jiayuan Song, Zhengmao Zhu, Wenhu Chen, Pengyu Zhao, and Junxian He. Webexplorer: Explore and evolve for training long-horizon web agents, 2025. URL https://arxiv.org/abs/2509.06501.

Rui Lu, Zhenyu Hou, Zihan Wang, Hanchen Zhang, Xiao Liu, Yujiang Li, Shi Feng, Jie Tang, and Yuxiao Dong. Deepdive: Advancing deep search agents with knowledge graphs and multi-turn rl, 2025. URL https://arxiv.org/abs/2509.10446.

Lang Mei, Xiaohan Yu, Chong Chen, Liyan Liu, Xiangnan Chen, Jinchao Ma, Chao Feng, Li Huang, Siyu Mo, Sichen Kang, Yunkun Xu, Zhihan Yang, Zhujun Xue, Jingren Zhang, Qing He, Yingdi Huang, Hao Jiang, Ziao Ma, Zewei Pan, Minhao Sun, Zhuo Tao, Jinzhao Xiao, Gangtao Xin, Huanyao Zhang, Wenjian Zhang, Jiangshan Zhang, Guojie Zhu, Fangzhou Zou, Jiaxin Mao, and Wentao Zhang. Searchart: Training long-horizon search agent with scalable synthetic and verified task, 2026. URL https://arxiv.org/abs/2607.24850.

MiroMind Team et al. MiroThinker-1.7 & H1: Towards heavy-duty research agents via verification, 2026. URL https://arxiv.org/abs/2603.15726v1.

Debayan Mukhopadhyay, Utshab Kumar Ghosh, and Shubham Chatterjee. Bridge evidence: Static retrieval utility does not predict causal utility in multi-step agentic search, 2026. URL https://arxiv.org/abs/2607.15253.

Fabio Petroni, Aleksandra Piktus, Angela Fan, Patrick Lewis, Majid Yazdani, Nicola De Cao, James Thorne, Yacine Jernite, Vladimir Karpukhin, Jean Maillard, Vassilis Plachouras, Tim Rocktäschel, and Sebastian Riedel. KILT: a benchmark for knowledge intensive language tasks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2523–2544, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.200. URL https://aclanthology.org/2021.naacl-main.200/.

Thinh Pham, Nguyen Nguyen, Pratibha Zunjare, Weiyuan Chen, Yu-Min Tseng, and Tu Vu. SealQA: Raising the bar for reasoning in search-augmented language models, 2026. URL https://arxiv.org/abs/2506.01062.

Long Phan et al. Humanity's last exam, 2026. URL https://arxiv.org/abs/2501.14249.

Zile Qiao, Guoxin Chen, Xuanzhong Chen, Donglei Yu, Wenbiao Yin, Xinyu Wang, Zhen Zhang, Baixuan Li, Huifeng Yin, Kuan Li, Rui Min, Minpeng Liao, Yong Jiang, Pengjun Xie, Fei Huang, and Jingren Zhou. Webresearcher: Unleashing unbounded reasoning capability in long-horizon agents, 2025. URL https://arxiv.org/abs/2509.13309.

Jiacheng Sang, Mengyuan Li, Sanxing Chen, Yukun Huang, Yu Feng, and Bhuwan Dhingra. Searchatlas: Analyzing agentic search strategies via evidential query graphs, 2026. URL https://arxiv.org/abs/2609.10901.

Alon Talmor and Jonathan Berant. The web as a knowledge base for answering complex questions, 2018. URL https://arxiv.org/abs/1803.06643.

Zhengwei Tao, Jialong Wu, Wenbiao Yin, Junkai Zhang, Baixuan Li, Haiyang Shen, Kuan Li, Liwen Zhang, Xinyu Wang, Yong Jiang, Pengjun Xie, Fei Huang, and Jingren Zhou. Webshaper: Agentically data synthesizing via information-seeking formalization, 2025. URL https://arxiv.org/abs/2507.15061.

Tongyi DeepResearch Team et al. Tongyi deepresearch technical report, 2026. URL https://arxiv.org/abs/2510.24701.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Musique: Multi-hop questions via single-hop question composition, 2022. URL https://arxiv.org/abs/2108.00573.

Jason Wei, Zhiqing Sun, Spencer Papay, Scott McKinney, Jeffrey Han, Isa Fulford, Hyung Won Chung, Alex Tachard Passos, William Fedus, and Amelia Glaese. Browsecomp: A simple yet challenging benchmark for browsing agents, 2025. URL https://arxiv.org/abs/2504.12516.

Jialong Wu, Baixuan Li, Runnan Fang, Wenbiao Yin, Liwen Zhang, Zhengwei Tao, Dingchu Zhang, Zekun Xi, Gang Fu, Yong Jiang, Pengjun Xie, Fei Huang, and Jingren Zhou. Webdancer: Towards autonomous information seeking agency, 2025. URL https://arxiv.org/abs/2505.22648.

Jian Xie, Tianhe Lin, Zilu Wang, Yuting Ning, Yuekun Yao, Tianci Xue, Zhehao Zhang, Zhongyang Li, Kai Zhang, Yufan Wu, Shijie Chen, Boyu Gou, Mingzhe Han, Yifei Wang, Vint Lee, Xinpeng Wei, Xiangjun Wang, Yu Su, and Huan Sun. QUEST: Training frontier deep research agents with fully synthetic tasks, 2026. URL https://arxiv.org/abs/2605.24218.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering, 2018. URL https://arxiv.org/abs/1809.09600.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models, 2023. URL https://arxiv.org/abs/2210.03629.

Yuqi Zhu, Jintian Zhang, Zhenjie Wan, Yujie Luo, Shuofei Qiao, Zhengke Gui, Da Zheng, Lei Liang, Huajun Chen, and Ningyu Zhang. LightThinker++: From reasoning compression to memory management, 2026. URL https://arxiv.org/abs/2604.03679.

A MPF Scoring and RCD Properties

A.1 Feature construction

For a grounded fact e, let p_e be its normalized predicate, κ_e its fact-type label, and (v_i, ρ_i, τ_i) its participants. The selector uses

$$\sigma(e) = (p_e, \kappa_e, \mathrm{sort}[(\rho_i, \tau_i)]_{i=1}^{a}), \tag{14}$$

$$\Gamma(e) = \{(p_e, \rho_i, \tau_i, v_i)\}_{i=1}^{a}. \tag{15}$$

For branch o, $Z^o_{\mathrm{sch},t}$ contains its $\sigma(e)$ values and $Z^o_{\mathrm{con},t}$ contains the union of its $\Gamma(e)$ values. $Z^o_{\mathrm{doc},t}$ contains non-root documents retrieved, expanded, or cited as evidence. $Z^o_{\mathrm{fact},t}$ contains stable identifiers for new facts and for the source facts used to issue the branch's acquisition actions. All four sets are reconstructed from the recorded branch trace; selection makes no additional model call.

Exact ties in Equation 10 follow the fixed subset order induced by assigning BFS, DFS, and SA-RW to bit positions 0, 1, and 2.

A.2 Why MPF favors balanced, nonredundant evidence

Coverage alone favors larger subsets even when their branches repeatedly acquire the same evidence. Within each channel, $r_{k,t}$ combines retained coverage with branch efficiency, so repeated features reduce channel quality. After max–min balance and mean quality have been matched, n_t aggregates efficiency across the four channels as a later tie-breaker. This hierarchy rewards complementary evidence while preserving balance between semantic structure and evidence acquisition.

Proposition 1 (MPF selection behavior)**.** *For any policy subset A,*

$$\min\{s_t(A), v_t(A)\} = \min_{\lambda \in [0,1]} [\lambda s_t(A) + (1-\lambda) v_t(A)]. \tag{16}$$

For a feature channel with a universe of $N > 0$ items, suppose A covers u unique items across $\ell > 0$ total branch items. Adding a branch containing b items, of which q are new, changes its channel score by

$$\Delta r = \frac{1}{2}\left[\frac{q}{N} + \frac{\ell q - ub}{\ell(\ell + b)}\right]. \tag{17}$$

If $q = 0$ and $u, b > 0$, then $\Delta r < 0$. Consequently, a branch that adds no new feature in any channel and does not increase realized depth cannot outrank the subset without that branch under Equation 10.

Proof. The expression in Equation 16 is affine in λ, so its minimum on $[0, 1]$ is attained at an endpoint. For the second result, the old channel score is $\frac{1}{2}(u/N + u/\ell)$ and the new score is $\frac{1}{2}((u + q)/N + (u + q)/(\ell + b))$; subtraction gives Equation 17. When $q = 0$, the change is $-ub/[2\ell(\ell + b)]$. If this holds in every channel and depth does not increase, the semantic, evidence, and nonredundancy coordinates cannot improve, while the cardinality tie-breaker strictly favors the smaller subset. □

A.3 RCD AND THE COMMITTED-STATE INVARIANT

For each policy o, let $\mathcal{F}_t^o$ be the frontier it produces after removing entities already expanded in the updated observed state, and let x_t^o be its first entity, or $\perp$ if the frontier is empty. Let $N_t^o \subseteq \mathcal{F}_t^o$ contain frontier entities absent from the current committed entity hypergraph, let ℓ_t^o be realized depth, and let β_t^o be the bottleneck score of singleton policy subset $\{o\}$. ACTS selects

$$o_t^* = \operatorname*{lexmax}_{o \in \mathcal{O}} \left(\mathbb{I}[x_t^o \neq \perp], \mathbb{I}[x_t^o \neq \perp \wedge x_t^o \notin V_{\mathrm{E},t}^{\mathrm{com}}], \frac{|N_t^o|}{|\mathcal{F}_t^o|}, \right.$$
$$\left. \frac{|\mathcal{F}_t^o|}{\max_{j \in \mathcal{O}} |\mathcal{F}_t^j|}, \min\{\ell_t^o / H, 1\}, \beta_t^o, -\operatorname{rank}(o) \right), \tag{18}$$

where undefined ratios are zero and $\operatorname{rank}(o)$ is the fixed policy order. If another round is scheduled, ACTS sets $x_{t+1} = x_t^{o_t^*}$.

The RCD block in Figure 2 summarizes the separation between retention and continuation.

For a selected continuation entity x, let $\mathrm{Supp}_{\mathrm{obs}}(x)$ contain all of its recorded support: the acquisition actions leading to it, their source facts and retrieved documents, and all grounded facts extracted from those documents. The records added to the committed state before the next round are

$$\mathcal{K}_{t+1} = \mathrm{Supp}_{\mathrm{obs}}(x_{t+1}) \setminus \mathcal{G}_t^{\mathrm{com}}. \tag{19}$$

Proposition 2 (Committed-state invariant)**.** *After every completed round, the committed entity and document-navigation hypergraphs are subgraphs of their observed counterparts. Every committed fact retains its source document, and every committed document required by a retained fact or selected continuation entity is reachable from d_0 through committed document-navigation hyperedges. Adding $\mathcal{K}_{t+1}$ requires no new retrieval.*

Proof. At initialization, both states contain the same root document and facts grounded in it. During a round, Equation 5 adds every grounded branch increment to the observed state. Equation 11 copies selected increments and previously observed support records into the committed state, so the subset relation is preserved. Each selected fact is copied with its source-document identifier, and each copied acquisition action connects an already reachable source document to its retrieved documents. If continuation comes from an uncommitted branch, Equation 19 copies all of its recorded support before that continuation entity is used. All added records therefore already exist in the observed state, establishing rooted reachability without additional retrieval. □

A.4 EXPANSION-STEP BOUND

With three policies, ACTS scores $2^3 - 1 = 7$ nonempty policy subsets per round. The full configuration executes at most three expansion steps per policy for three rounds, giving 27 policy expansion steps. Cached results can reduce the number of retrieval requests but do not change this bound.

Algorithm 1: Multi-round evidence-acquisition topology construction

Input: resolved root $r = (d_0, v_0)$, representation ϱ, search policies $\mathcal{O}$, admissible policy subsets $\mathfrak{A}$, budget schedule $\mathbf{b}$, continuation rule Cont.
Initialize: extract facts grounded in d_0; initialize the observed and committed states from d_0, v_0, and these facts; set $\mathcal{K}_1 = \emptyset$.
For $t = 1, \ldots, T$: (1) fork every $o \in \mathcal{O}$ from the same observed state and active entity; (2) execute the policy expansion budget $H_{t,o}$; (3) union all branch increments into the observed state; (4) score every $A \in \mathfrak{A}$ and commit $\mathcal{K}_t$ plus the increments in A_t^*; (5) apply Cont to choose a continuation entity from the policy frontiers using Equation 18; (6) if another round will execute, add its required recorded support as $\mathcal{K}_{t+1}$; stop if no continuation entity remains.
Output: committed state, writer input, observed state, branch traces, and acquisition provenance.

Figure 3: The parameterized ACTS constructor. LHR records branch increments, MPF scores the admissible policy subsets and fuses the selected increments, and RCD separates evidence retention from continuation. The experimental configuration uses three policies, all seven nonempty subsets, and $H = T = 3$; every branch executes retrieval against the corpus.

B PIPELINE AND MEASUREMENT SPECIFICATIONS

B.1 GENERATION ROLES AND SHARED CONSTRUCTION SETTINGS

Table 5 separates models that construct questions from models that test or solve them. The same 640 root pairs $r = (d_0, v_0)$ are assigned to every arm. Each arm uses the same KILT snapshot, $H = 3$, $T = 3$, and at most 27 policy expansion steps.

The ordinary-graph SA-RW arm exposes grounded binary projections of the same branch trace whose parent n-ary facts are preserved by the hypergraph SA-RW arm. The all-branch union arm commits all three branch increments. The entity-only selector removes the document channel from the evidence score, while the full ACTS selector uses the objective in Equation 10. The size-matched random control retains a deterministically seeded subset whose cardinality matches the full ACTS selector's choice.

For downstream evaluation, every arm contributes 2,000 complete, judged-correct teacher trajectories to the same SFT protocol and is evaluated on the same BC-130 split.

Table 5: Model roles in the six-arm construction ablation. Thinking is disabled for all construction and no-tool calls. The trajectory teacher uses its native reasoning mode and receives only the question and tools.

Role	Model	Inputs and information access
Long-tail entity proposals	GPT-5.5	Domain and proposal prompt; frozen before the construction ablation
Grounded-fact extraction	Kimi-K3	Retrieved KILT documents; no downstream outcomes
Writer and editors	Kimi-K3	Committed entity hypergraph, candidate answer, and its construction path; no uncommitted branches or selector scores
No-tool solver	GLM-5.1	Question only; no tools, graph, evidence, or reference answer
No-tool answer judge	GPT-5.5	Question, solver prediction, and reference answer
Trajectory teacher	Kimi-K3	Question and the online search, open, find, and Python tools only
Trajectory answer judge	GPT-5.5	Question, extracted prediction, and reference answer

Table 6: Training recipes for the reported end-to-end checkpoints. The 11,651-trajectory training set is shared across backbones; Qwen3-30B-A3B also uses a multi-teacher 16,792-trajectory recipe, marked with $\star$.

Component	Qwen3-30B-A3B (11,651 / 16,792*)	Qwen3.8-27B	Qwen3.5-35B-A3B
Student	Qwen3-30B-A3B-Thinking-2507	Qwen3.8-27B	Qwen3.5-35B-A3B
Teacher trajectories	11,651 Kimi-K3 / 8,077 Kimi-K3 + 8,715 GLM-5.1	11,651 Kimi-K3	11,651 Kimi-K3
Unique QA	11,651 / 13,171	11,651	11,651
Optimization	SFT; 6 epochs; cosine LR; peak 5×10^{-5}; global batch 128	SFT; 6 epochs; cosine-to-min LR; peak 5×10^{-6}; minimum 5×10^{-7}; global batch 32	SFT; 6 epochs; cosine LR; peak 1×10^{-5}; global batch 32
Sequence handling	262,144 tokens; packing	262,144 tokens; packing; Qwen3.8 native template; Hermes agent rendering	262,144 tokens; packing; Qwen3.5 native template; Hermes agent rendering
Training endpoint	epoch 6 / epoch 6	epoch 6	epoch 6
Evaluation judge	GPT-5.5 strict answer equivalence	GPT-5.5 strict answer equivalence	GPT-5.5 strict answer equivalence

B.2 Training-data curation

The 11,651-trajectory ACTS training set is selected by a fixed curation policy. It requires a judged-correct teacher trajectory, an extractable nonempty final prediction, an acceptable terminal status, no benchmark-URL leakage, no recognized tool-error response, and a unique normalized question. It then applies the fixed caps in Table 7. Filtering selects complete trajectories without editing their messages.

Table 7: Implemented behavior and size caps for the 11,651-trajectory training set.

Quantity	Maximum	Quantity	Maximum
Assistant turns	100	Total tool calls	220
Search / Open calls	120 / 90	Find / Python calls	50 / 8
Search per assistant turn	1.6	Tools per assistant turn	3.6
Total message characters	420K	Tool-response characters	360K
Single tool response	80K	Reasoning characters	80K

B.3 Multi-teacher supervision for Qwen3

The starred Qwen3 recipe combines 8,077 curated Kimi-K3 trajectories with 8,715 curated GLM-5.1 trajectories, yielding 16,792 trajectories over 13,171 unique questions. The pools overlap on 3,621 questions, for which the recipe retains independently generated trajectories from both teachers. This design increases task coverage while exposing the student to complementary search behaviors on shared tasks.

We use multi-teacher supervision after observing a substantial difference in how readily Qwen3-30B fits trajectories from the two teachers. Figure 4 compares token-level negative log-likelihood (NLL) across single- and multi-teacher SFT. The matched-protocol Kimi-11K, GLM-5.1-only, and multi-teacher runs use the same backbone, six-epoch protocol, cosine schedule with peak learning rate 5×10^{-5}, global batch 128, 262,144-token context, and sequence packing. The figure also shows the source-matched 8,077-trajectory Kimi run under its original three-epoch recipe. The mean/final NLL is 0.935/0.775 for Kimi-11K SFT, 0.460/0.400 for GLM-5.1-only SFT, and 0.529/0.457 for multi-teacher SFT. Relative to Kimi-11K SFT, multi-teacher SFT also raises full BrowseComp accuracy from 43.76% to 50.47% under anti-contamination evaluation. The matched-protocol single-teacher contrast is consistent with a teacher–student capacity gap, while the mixed result suggests that the more readily learned GLM-5.1 trajectories provide complementary supervision that helps Qwen3 benefit from a training set retaining Kimi-K3 trajectories. The remaining curves place this pattern in cross-backbone context: under their respective training recipes, Kimi-only final NLL is 0.775 for Qwen3, 0.595 for Qwen3.5, and 0.617 for Qwen3.8.

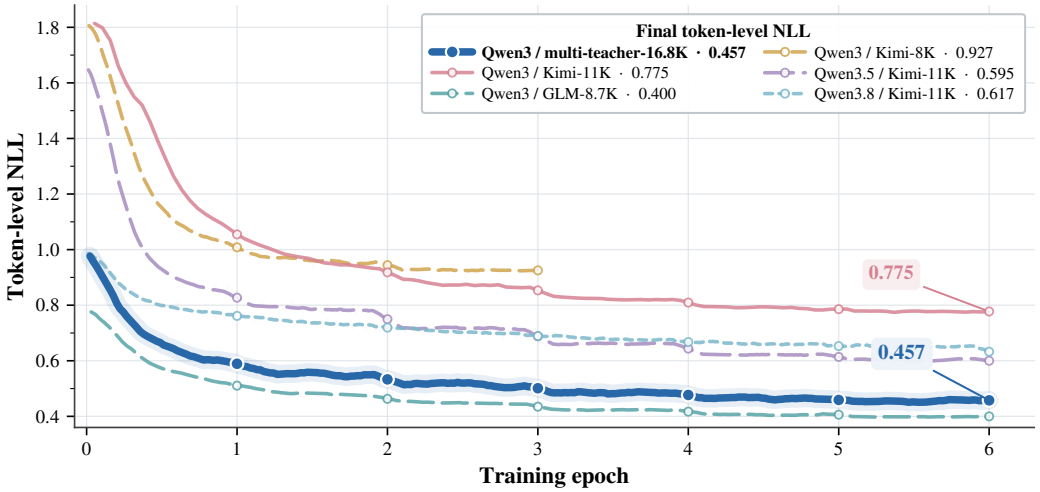


Figure 4: Token-level NLL under single- and multi-teacher SFT. Faint lines show step-level losses, emphasized curves show smoothed trends, and markers denote completed epochs. Under the same Qwen3 training protocol, GLM-5.1-only and multi-teacher SFT converge below Kimi-11K SFT; the dashed Kimi-8K curve provides a source-matched Kimi-only reference under the original three-epoch recipe. The Kimi-11K curves compare student backbones trained on the same 11,651 trajectories.

B.4 MPF BEHAVIOR AND SENSITIVITY

For channel k in round t, we measure overlap across the full three-policy candidate pool as

$$\text{Overlap}_{k,t} = 1 - \frac{\left|\bigcup_{o\in\mathcal{O}} Z^o_{k,t}\right|}{\sum_{o\in\mathcal{O}} |Z^o_{k,t}|}. \tag{20}$$

This is the complement of the efficiency term used by the selector when $A = \mathcal{O}$. We average the four relation-pattern, entity-role, document, and fact channels; every channel is nonempty in the analyzed decisions. Table 8 shows that the increase occurs in every channel.

Table 8: Cross-branch overlap and singleton retention by construction round (%).

Statistic	Round 1	Round 2	Round 3
Relation-pattern overlap	41.41	58.21	57.87
Entity-role overlap	41.85	58.50	58.00
Document overlap	35.70	55.05	54.23
Fact overlap	41.04	58.18	57.66
Mean cross-branch overlap	40.00	57.49	56.94
Singleton retention	28.98	80.45	80.62

For constructions that reach both rounds, mean cross-branch overlap increases by 17.92 points from round 1 to 2 and by 4.18 points from round 2 to 3. Thus the small change between the unpaired round-2 and round-3 aggregates reflects cohort composition rather than a reversal of the within-construction trend.

Table 4 weights each construction decision by the number of curated training trajectories derived from that construction, matching its frequency in the SFT data. Counting each of the 3,566 constructions with complete selector records once produces 9,291 selection rounds and the distribution in Table 9. The ordering of all seven subsets is unchanged, and no share moves by more than 0.74 percentage points. The unique-construction cross-branch overlap means are 40.37%, 57.14%, and 56.54% in rounds 1–3; the paired increases are 17.59 and 3.97 points. RCD selects a continuation

policy outside the retained policy subset in 2,096/5,725 transitions (36.61%), compared with 36.75% under training-trajectory weighting.

Table 9: Selector distribution under training-trajectory and unique-construction weighting (%). The analysis uses trajectories with complete scores for all seven candidate policy subsets.

Retained policy subset	Training-trajectory weighted	Unique-construction weighted
BFS	36.29	35.55
DFS+SA-RW	18.65	19.17
DFS	14.40	14.83
BFS+SA-RW	10.88	10.78
SA-RW	10.14	10.13
BFS+DFS+SA-RW	5.39	5.47
BFS+DFS	4.25	4.07

B.5 Construction-path alignment

For trajectory i, let P_i^- be its ordered construction path after removing the answer node, B_i^- its non-answer bridge terms, and D_i^- its non-answer source documents. A path node is represented in the final question when the question contains its normalized entity name, its recorded obfuscation, or a grounded clue sharing at least three informative terms with the node evidence and attaining token-cosine similarity of at least 0.30. The question-side path coverage is the macro-average

$$F_Q = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P_i^-|}\sum_{v\in P_i^-}\mathbb{I}[v \text{ is represented in } q_i]. \tag{21}$$

In the trajectory teacher's rollout, a bridge is recovered when it appears in a search query. A path node is recovered when the teacher searches its entity, opens a page whose title names it, or opens one of its recorded source documents. Source-document matching normalizes URL encoding, Wikipedia titles, fragments, and query strings. We also report the fraction of trajectories that open at least one document in D_i^- and the longest-common-subsequence (LCS) recall between P_i^- and the nodes' first-observation order.

The two controls preserve the unchanged $(q_i, \text{trajectory}_i)$ pair while replacing its associated construction through a circular half-shift or a seeded random derangement. They preserve model behavior, question difficulty, and tool budget while breaking construction-path alignment.

Table 10: Alignment between construction paths, final questions, and teacher search across all 11,651 curated trajectories. Metrics are percentages; Eligible is a count. Each metric excludes the answer node and its source documents; source opening is defined when at least one non-answer source URL is recorded.

Signal	Aligned	Half-shift	Random	Eligible
Path-node coverage in the question	62.78	0.76	0.83	11,651
$\geq$ 2 path nodes represented	82.52	0.80	0.91	11,651
Construction bridge terms searched	51.04	0.69	0.67	11,624
Path nodes recovered by the teacher	59.95	0.71	0.75	11,651
$\geq$ 1 construction source opened	66.37	0.24	0.31	6,741
Path-order LCS recall	41.52	0.68	0.69	11,651

All 11,651 trajectories join exactly to their recorded synthesized QA, and every teacher user message matches the final question. The online rollouts issue 174,670 Open calls, of which 165,808 succeed (94.93%), and 11,610 trajectories open at least one online document. Varying the grounded-clue threshold from 0.25 to 0.35 changes aligned path-node coverage from 67.09% to 57.88%; the corresponding random-control values remain between 1.03% and 0.67%.

B.6 ACTS–QUEST COMPARISON DETAILS

We compare ACTS with QUEST (Xie et al., 2026) at a similar training-trajectory count. Both recipes use Kimi-K3 rollouts, six-epoch SFT of Qwen3.8-27B, and the same online evaluation on the QUEST BC-130 split and larger BC-200 diagnostic.

The comparison uses 4,670 ACTS trajectories and 4,516 QUEST trajectories. On the BC-200 diagnostic, the full 11,651-trajectory ACTS checkpoint answers 152/200 questions (76.0%); its headline result remains the complete BrowseComp evaluation in Table 1.

Both teacher rollouts use Kimi-K3 with temperature 1.0, nucleus probability 0.95, a 32,768-token per-response output cap, at most 500 model loops, and the registered `browser.search/open/find` and `python` tools. Table 11 summarizes the shared rollout, training, and evaluation configuration.

Table 11: Training and evaluation settings for the matched ACTS–QUEST data-recipe comparison.

Component	ACTS	QUEST
Question population	4,670 ACTS questions	4,670 released objective questions (Xie et al., 2026)
Teacher and tools	Kimi-K3; search, open, find, Python	Kimi-K3; search, open, find, Python
Teacher rollout decoding	temperature 1.0; top-p 0.95; 32,768-token output cap; 500 loops	temperature 1.0; top-p 0.95; 32,768-token output cap; 500 loops
Training trajectories / tool calls	4,670 / 243,613	4,516 / 135,898
Student training	Qwen3.8-27B; packed 262,144-token sequences; peak/min LR $5\times10^{-6}/5\times10^{-7}$; global batch 32	Same
Training epochs	6	6
Student evaluation	Qwen3.8-27B; 32,768-token output cap; QUEST BC-130 and BC-200 diagnostic; online four tools; GPT-5.5 judge	Same
QUEST BC-130 result	89/130 (68.46%)	78/130 (60.00%)
BC-200 diagnostic result	130/200 (65.00%)	103/200 (51.50%)

C DETAILED MEASUREMENT PROTOCOL

C.1 BENCHMARK PROTOCOLS

Our online evaluation covers all 1,266 BrowseComp questions, the 100-question xBench-DeepSearch-2510 set, all 111 SEAL-0 questions, and all 2,158 HLE-Text questions. Each uses search, page opening, in-page finding, and sandboxed computation. The agent receives no construction-time hypergraphs, reference answer, or supporting documents. The profile permits at most 500 tool calls within 501 model rounds.

Because public benchmark artifacts can be indexed by Web search, the anti-contamination protocol used for all ReAct and ACTS results in Table 1 blocks `huggingface.co` at both the search and page-opening layers. The remaining evaluation settings follow the protocol above.

Within each benchmark, GPT-5.5 receives the question, prediction, and frozen reference answer under the corresponding frozen answer-equivalence prompt and returns one judgment for every benchmark item. We report

$$\text{Accuracy} = \frac{N_{\text{judged correct}}}{N_{\text{benchmark questions}}}, \tag{22}$$

with every benchmark question in the denominator.

C.2 Six-arm construction ablation measurements

C.2.1 Treewidth diagnostic

For an entity hypergraph G, let $P(G)$ be its primal projection: two entities are adjacent when they participate in the same entity hyperedge. We report

$$\mathbb{I}[\hat{w}(G) \geq 4], \tag{23}$$

where $\hat{w}(G)$ is the treewidth of $P(G)$ computed exactly for at most 12 vertices and otherwise estimated by the smaller width returned by min-degree and min-fill elimination. We use this quantity as a relative treewidth diagnostic across the six ablation arms.

C.2.2 Outcome and tool-use metrics

For each arm, the tool-required rate is

$$\text{ToolRequired} = \frac{N_{\text{valid no-tool decisions}} - N_{\text{solved without tools}}}{N_{\text{valid no-tool decisions}}}. \tag{24}$$

For the six-arm construction ablation, teacher solve rate is computed separately from benchmark accuracy:

$$\text{TeacherSolve}_a = \frac{N_{\text{judged correct},a}}{N_{\text{complete teacher trajectories},a}}, \tag{25}$$

where the denominator is the fixed 1,800-trajectory sample for arm a. GPT-5.5 is the answer judge for this analysis; the trajectories themselves were produced with the same Kimi-K3 rollout settings.

For a trajectory, Calls counts the structured tool calls issued across all teacher messages. Sources counts the distinct source URLs recorded on nodes and hyperedges in the committed entity hypergraph supplied to the writer. The main ablation table reports both means over judged-correct teacher trajectories.

Every trajectory is linked to its root pair, acquisition actions, committed state, writer record, filter decisions, no-tool request, teacher trajectory, final prediction, answer judgment, and training membership. Counts are reported separately for generated items, items without infrastructure failures, judged items, and training examples. Comparisons use frozen question denominators and the same training and evaluation protocol across arms.

Table 12: Tool-use diagnostics for the six ablation arms. $\text{Calls}_{\text{all}}$ uses all 1,800 trajectories per arm; $\text{Calls}_{\text{corr}}$ uses only judged-correct teacher trajectories. Open/search is a call-count ratio. Open success is the share of Open calls returning content; Find execution includes both valid match and no-match responses, while Find match is the share of all Find calls that return a match.

Arm	$\text{Calls}_{\text{all}}$	$\text{Calls}_{\text{corr}}$	Open/search	Open success	Find execution	Find match
Ordinary-graph SA-RW	29.07	27.47	0.522	90.52	91.91	57.29
Hypergraph SA-RW	30.73	28.71	0.509	91.31	90.05	58.81
All-branch union	33.63	31.10	0.500	91.76	93.04	58.16
Entity-only selector	33.18	30.69	0.476	91.09	92.71	57.53
ACTS (full)	32.99	31.60	0.504	91.46	92.52	57.00
Size-matched random	32.30	30.04	0.496	91.30	93.19	56.90

D Complete Recorded Example: From a Long-Tail Entity to a Teacher Trajectory

This section traces one recorded ACTS example through the artifacts produced at each stage. The root is *Uprising of Ivaylo*; the generated answer is *Bulgarians*. We selected this generated example because its construction record, question revisions, no-tool gate, complete 85-call teacher rollout, and answer judgment can be joined by the same task key. The case illustrates the pipeline artifact chain; dataset curation is reported separately in Appendix B.2.

Artifact chain. Long-tail proposal → KILT root resolution → root fact extraction → three-round acquisition → committed state → writer evidence packet → question edits → no-tool gate → independent teacher trajectory → answer judgment.

D.1 Root Proposal, Resolution, and Graph Initialization

The long-tail proposer emits an entity record before any graph construction. The recorded fields for this case are:

Long-tail proposal record

`name`	*Uprising of Ivaylo*
`domain`	peasant rebellion
`description`	A late thirteenth-century Bulgarian peasant and military revolt led by Ivaylo against royal and Mongol pressures.
Disambiguation hint	`Ivaylo Bulgaria 1277 1280`

The resolver queries KILT with `Uprising of Ivaylo`. The rank-1 result is an exact normalized-title match to `kilt://Uprising of Ivaylo`. Its retrieved opening span states that the uprising was a rebellion of the Bulgarian peasantry against Emperor Constantine Tikh and the Bulgarian nobility. A grounded extractor turns this single span into the following four initial entity hyperedges:

Initial entity hypergraph records

Predicate	Participants	Grounding
`was a rebellion against`	Uprising of Ivaylo; Constantine Tikh	"rebellion of the Bulgarian peasantry against the incompetent rule of Emperor Constantine Tikh and the Bulgarian nobility"
`was a rebellion against`	Uprising of Ivaylo; Bulgarians	same retrieved span; `Bulgarians` is bound as the target group
`was a rebellion of`	Uprising of Ivaylo; Bulgarians	same retrieved span; `Bulgarians` is bound as the rebel group
`was a rebellion against the incompetent rule of`	Uprising of Ivaylo; Constantine Tikh	same retrieved span; `Constantine Tikh` is bound as the ruler

These four initial edges are binary instances in an entity hypergraph that also supports higher-order relations. They provide grounded frontier entities—most importantly *Bulgarians* and *Constantine Tikh*—from which the three policies can take their first distinct navigation actions. A higher-order edge appears after the first acquisition round and later becomes the question's answer anchor.

D.2 Three-Round Acquisition and Selection

Each round starts from the shared observed state. BFS, DFS, and self-avoiding random walk (SA-RW) are each budgeted for up to three logical searches, and every returned fact is attached to the query and KILT document that made it reachable. MPF evaluates all seven nonempty policy subsets; RCD then selects the next construction frontier independently of the retained subset.

Recorded construction trace

Round	Active entity	Queries issued by the branches	Retained subset	Next frontier
1	*Uprising of Ivaylo*	BFS/DFS: `constantine tikh of bulgaria` → `bulgarians` → `macedonians (ethnic group)`; SA-RW: `bulgarians` → `y-dna`	BFS, 0.889	BFS → *Romanians*
2	*Romanians*	BFS/DFS: `romanians` → `balkans` → `transylvania`; SA-RW: `romanians` → `transylvania` → `kingdom of romania`	SA-RW, 0.942	**BFS** → *Romanian Revolution*
3	*Romanian Revolution*	BFS: `romanian revolution` → `socialist republic of romania` → `nicolae ceausescu`; DFS ends at `communist state`; SA-RW ends at `communist`	DFS, 0.853	DFS → *Juche* (horizon reached)

Round 2 makes both selection mechanisms concrete. The seven recorded MPF scores are shown below. The table reports (s_t, v_t, n_t). From these values, MPF forms the stored objective tuple $(\min\{s_t, v_t\}, (s_t + v_t)/2, n_t, -|A|)$ and appends a deterministic subset-order key. MPF compares the full tuple lexicographically.

Round 2 candidate subsets

Subset	s_t	v_t	n_t	Subset	s_t	v_t	n_t
BFS	0.736	0.846	1.000	BFS+SA-RW	0.883	0.817	0.642
DFS	0.736	0.846	1.000	DFS+SA-RW	0.883	0.817	0.642
BFS+DFS	0.569	0.596	0.500	BFS+DFS+SA-RW	0.827	0.720	0.461
SA-RW	**1.000**	**0.942**	**1.000**				

The serialized fields `reasoning_value` and `evidence_value` correspond to s_t and v_t. The corresponding recorded decision is:

Round 2: retained evidence and continuation

```
''retained_subset'': {''operators'': [''self_avoiding_random_walk''],
  ''reasoning_value'': 1.000, ''evidence_value'': 0.942,
  ''objective'': [0.942, 0.971, 1.000, -1, -4]}
''continuation'': {''selected_operator'': ''breadth_first_search'',
  ''active_tip_entity_name'': ''romanian revolution'',
''scheduled_for_next_round'': true}
```

Thus, SA-RW supplies the evidence committed after round 2, but round 3 begins from the BFS frontier at *Romanian Revolution*. This is the recorded RCD event: retention answers which evidence should be exposed downstream, whereas continuation answers where acquisition should proceed next.

Across all branches, the observed state contains 37 entities, 39 entity hyperedges, 42 documents, and 26 document-navigation hyperedges. Selective retention produces a committed state with 36 entities, 37 entity hyperedges, 28 documents, and 11 document-navigation hyperedges. The key higher-order record is:

Committed answer-anchor hyperedge

```
predicate: situates_in_cluster_with
participants: [Bulgarians, Romanians, Macedonians, Gagauzes]
source: kilt://Bulgarians
```

Evidence span: "the whole genome SNP data situates Bulgarians in a cluster with Romanians, Macedonians and Gagauzes."

The four-way relation supplies the answer anchor. Other committed edges make each non-answer participant describable through independently retrievable relations, rather than naming it directly in the question.

D.3 Writer Input, Question Realization, and Screening

The writer packet contains the complete committed entity view: 36 entity nodes and 37 entity hyperedges represented as grounded facts from nine KILT source documents. It includes neither policy identities nor MPF scores nor the construction search log. Five records used visibly in the realized question are listed below.

Selected records from the writer evidence packet

Bulgarians. `situates_in_cluster_with` → Romanians, Macedonians, and Gagauzes. The evidence span places the four groups in one whole-genome SNP cluster.

Macedonians (ethnic group). `share_lineages_with` → Greeks and Romanians.

Kingdom of Romania. `was` → constitutional monarchy; `existed_until` → Michael I's 1947 abdication; `proclaimed_as` → socialist republic.

Socialist Republic of Romania. `was_under` → Marxist-Leninist one-party rule from 1947 to 1989.

Romanians; Communist state. `could_be_found_in` → Balkans; `adopted` → Juche.

The initial writer output contains direct concepts such as *constitutional monarchy*, *Balkan*, and *Juche*. The recorded graph-grounded editing stage replaces them as follows:

Recorded question edits

Direct concept	Relational clue in the final question
constitutional monarchy	"the governance system that was ended by the abdication of King Michael I of Romania"
Balkan	"the region where many Vlachs could be found"
Juche	"the ideology that was adopted by the state established in North Korea"

Final generated question

What ethnic group is situated in a genetic cluster alongside the inhabitants of the Eastern European country that was historically under the governance system that was ended by the abdication of King Michael I of Romania, a South Slavic people native to the region where many Vlachs could be found, and a Turkic group, where the country in question became a socialist republic under a Marxist-Leninist one-party system, and the South Slavic people share lineages with a Hellenic population native to a Mediterranean nation, which in turn is where a particular governance system that later adopted the ideology that was adopted by the state established in North Korea was established?

The stage-1, enhanced-question, and post-edit validators all return `ok: true`; the diversity selector retains the task. In the no-tool gate, the solver answers *Romanians*. The judge compares this prediction with *Bulgarians*, marks it incorrect, and therefore passes the question to teacher rollout.

No-tool gate record. `model_no_tool_answer:` "The ethnic group is the **Romanians**." `judge_response: incorrect gate_status: not_solved`

D.4 Native Excerpts from the Teacher Trajectory

The independent teacher receives the final question and the standard tool interface, not the construction-time hypergraphs. Its recorded trajectory contains 130 messages and 85 browser calls: 40 searches, 25 page opens, and 20 within-page finds. The excerpts below reproduce the original messages, tool arguments, and returned lines; ellipses mark omitted spans, citation markers are rendered as cursor–line ranges, and line wrapping is normalized.

Native excerpt A: decompose the question and form a hypothesis

Assistant reasoning (message 2). "We need solve convoluted question. Need identify ethnic group. Need browse sources and cite. Let's parse. [...] Eastern European country [...] Kingdom of Romania abolished Dec 30 1947, Romanian People's Republic. Inhabitants Romanians. [...] Asked ethnic group."

Tool call. `browser.search({''query'': ''genetic cluster Romanians Bulgarians Gagauz ethnic group cluster''})`

Tool response (message 3). `L11--L13: Genetic studies on Bulgarians [...] the southeastern group (Bulgarians and Macedonians) in a cluster with Romanians [...]`

`L14--L16: The Genetic Variability of Present-Day Bulgarians [...] Bulgarian samples form a cluster alongside Romanians, North Macedonians [...]`

Updated reasoning (message 4). "Search result likely genetic studies on Bulgarians says [...] Asked 'What ethnic group is situated in a genetic cluster alongside Romanians, Macedonians, and Gagauzes' [...] likely Bulgarians."

The search result is treated as a hypothesis rather than an answer. The teacher next verifies the non-answer entities encoded by the question. The following messages are reproduced from the middle of the same trajectory.

Native excerpt B: resolve the non-answer clue entities

Assistant reasoning and parallel tool calls (message 25). "Find 'South Slavic', 'lineages'."

`browser.find({''cursor'': 11, ''pattern'': ''South Slavic ethnic group''})`

`browser.find({''cursor'': 11, ''pattern'': ''lineages are also''})`

`browser.find({''cursor'': 11, ''pattern'': ''Vlachs''})`

Tool responses (messages 29, 32, and 37, excerpts).

`[18] L424--L425: Vlachs (Aromanians) [...] also inhabited this area in the Middle Ages [...]`

`[17] L1934--L1941: Such lineages are also typically found in [...] Greeks, Albanians, Romanians and Gagauzes.`

`[20] L220--L224: [a] South Slavic ethnic group native to the region of Macedonia in Southeast Europe.`

Parallel searches (message 58, selected calls).

`browser.search({''query'': ''King Michael abdication December 1947 Romanian monarchy ended People's Republic source Britannica''})`

`browser.search({''query'': ''Gagauz Turkic ethnic group Britannica''})`

Tool response (message 62, excerpt). `[39] L7--L9: The Gagauz are an Oghuz Turkic ethnic group native to southern Moldova and southwestern Ukraine.`

Later find/open sequence (messages 113–119).

`browser.find({''cursor'': 68, ''pattern'': ''forced abdication of King Michael''})`

`browser.open({''cursor'': 68, ''loc'': 280, ''num_lines'': 70})`

`[73] L297--L309: The final act was the forced abdication of King Michael and the proclamation of the Romanian People's Republic [...]. [The] 1948 [...] constitution [...] reserved ultimate authority for the party.`

Having resolved these clue entities, the teacher opens a population-genetics article and then the exact four-group source passage:

Native excerpt C: retrieve and inspect decisive evidence

Assistant reasoning (message 80). "Find cluster."

Tool call. `browser.find({''cursor'': 48, ''pattern'': ''form a cluster''})`

Tool response (message 81).

```
L0: match at L452

L1: form a cluster alongside Romanians, North Macedonians, and some
Greek

L2: individuals, thus supporting PCA and FST results. This
Eastern-Balkan cluster

L3: is then related with continental Greek and Western-Balkan population
clusters
```

Assistant reasoning (message 82). "Open lines."

Tool call. `browser.open({''cursor'': 48, ''loc'': 430, ''num_lines'': 55})`

Tool response (message 83, excerpt).

```
L451--L455: Herein, the Bulgarian samples form a cluster alongside
Romanians, North Macedonians, and some Greek individuals [...]. This
Eastern-Balkan cluster is then related with continental Greek and
Western-Balkan population clusters [...].
```

Updated reasoning (message 84). "This gives exact Bulgarian cluster with Romanians/North Macedonians [...]. The question likely draws from older summary."

Later tool call (message 86). `browser.open({''cursor'': 5, ''loc'': 590, ''num_lines'': 40})`

Tool response (message 87, excerpt).

```
L603--L609: the whole genome SNP data situates Bulgarians in a cluster
with Romanians, Macedonians and Gagauzes [...]
```

The teacher then checks the exact four-group relation on the *Bulgarians* page. The returned passage states that whole-genome SNP data places Bulgarians in a cluster with Romanians, Macedonians, and Gagauzes (messages 86–87). This matches the anchor relation used to construct the question, but it is reached through the teacher's own web search and page navigation.

Native excerpt D: final answer (message 129)

Answer: Bulgarians

"The clues resolve as follows:

1. **Romanians**—King Michael's forced abdication and the proclamation of the Romanian People's Republic both occurred on December 30, 1947; the subsequent constitution concentrated authority in the Communist Party. Thus the country is Romania. `[73:L297--L309]`

2. **Macedonians**—They are a South Slavic ethnic group native to Macedonia. Vlachs, or Aromanians, also inhabited that region. Their reported genetic lineages are additionally found among Greeks, matching the Hellenic clue. `[20:L220--L224; 18:L424--L425; 17:L1930--L1941]`

3. **Gagauz**—The genetic literature explicitly describes the Gagauz as Turkic-speaking. `[35:L522--L525]`

4. **The genetic-cluster clue**—Whole-genome SNP data are summarized as placing **Bulgarians** in a cluster with Romanians, Macedonians, and Gagauzes. A recent peer-reviewed genomic study likewise reports Bulgarian samples clustering with Romanians and North Macedonians. `[52:L603--L609; 50:L451--L455]`

Therefore, the ethnic group sought is the **Bulgarians**—meaning the modern South Slavic people, not the medieval Turkic Bulgars."

The stored answer is `bulgarians`; the extracted prediction is *Bulgarians*; and the answer judge records `is_correct: true`. This closes the recorded chain from long-tail proposal through acquisition, question writing, independent tool use, and answer judgment.

E Per-Example Provenance Records

To reconstruct every denominator in Section 4 while separating generation, infrastructure, judging, and training outcomes, the per-example schema records: (i) root-entity proposal, resolved root document, provenance, round, search policy, request, and cache keys; (ii) grounded fact, role, evidence span, source document, acquisition action, and graph-state IDs; (iii) retained policy subset, continuation entity, recorded support added to the committed state, writer input, QA revision, and filter decisions; and (iv) no-tool and teacher requests, trajectories, tool interactions, predictions, judgments, retries, infrastructure status, and SFT export ID.

F Limitations

The available corpus, grounded-fact extractor, and root distribution determine which construction paths ACTS can produce. The current constructor uses three policies and evaluates their seven nonempty policy subsets exactly, with at most $3HT$ policy expansion steps per root. Larger policy libraries require restricted candidate families or approximate subset selection.