ARGUS: EVIDENCE ASSEMBLY FOR SCALABLE DEEP RESEARCH AGENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep research agents answer complex questions by collecting and combining evidence from the web. Parallel search increases coverage, but independently generated trajectories can repeat the same findings while leaving important questions unresolved. Their growing histories also make evidence aggregation difficult. We propose **Argus**, which organizes search around a shared evidence graph. A Searcher executes independent ReAct rollouts, while a Navigator extracts source-linked findings and tentative claims, assesses support and contradictions, and dispatches follow-up searches for missing evidence. The Navigator then synthesizes an answer from the assembled graph. We train the Searcher and Navigator separately, using reinforcement learning to optimize the Navigator's search, verification, and synthesis decisions. The same trained Navigator supports one or multiple initial Searchers without retraining. With both components initialized from a 35B-A3B backbone, Argus improves answer accuracy over the standalone Searcher by 5.5 percentage points with one initial Searcher and 12.7 points with eight, averaged across eight benchmarks. At 64 initial Searchers, it reaches 86.2% on BrowseComp. In that configuration, 25.6M cumulative Searcher tokens are represented by an approximately 21.5K-token graph view for final synthesis. The graph lets final synthesis use evidence from a large search budget without reading all accumulated trajectories.

1 INTRODUCTION

Deep research agents answer questions that require repeated search and reasoning across sources (OpenAI, 2025b; Citron, 2024; Tongyi DeepResearch Team et al., 2025). Many use ReAct to alternate reasoning, tool calls, and observations (Yao et al., 2023b). Running longer trajectories or more agents increases search effort, but independent runs can repeat the same findings and miss the same constraint. A useful next query must address what the evidence is missing, not simply restart the original search.

Parallel test-time methods offer several ways to use additional compute, including self-consistency, candidate verification, and aggregation across trajectories (Wang et al., 2023; Snell et al., 2024; Li et al., 2025a; Lee et al., 2026; Zeng et al., 2026). A verifier can gather external evidence, while an aggregation agent can inspect trajectories on demand. We study a complementary way to organize this work: maintaining a shared record of claims, their supporting or conflicting evidence, and unresolved parts of the question. This record serves both query selection and final answer synthesis.

We propose **Argus**, a Searcher–Navigator architecture that treats deep research as evidence assembly. Each Searcher executes an independent ReAct rollout for a supplied query and returns its trace. The Navigator extracts retrieved findings and tentative claims into a shared evidence graph, recording source links and support or contradiction relations. It uses this state to assess under-supported claims, conflicting findings, and unanswered requirements, then requests targeted follow-up searches. Returned findings update the graph before the next decision. The loop ends when the Navigator stops or exhausts its budget.

After evidence assembly, the Navigator clears its working context and synthesizes an answer from the original question and the graph. The synthesis input contains source-linked findings and their relations, rather than the full Searcher histories. This reduces the context needed for the final pass; graph construction and verification still require the Navigator to process search returns.

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.

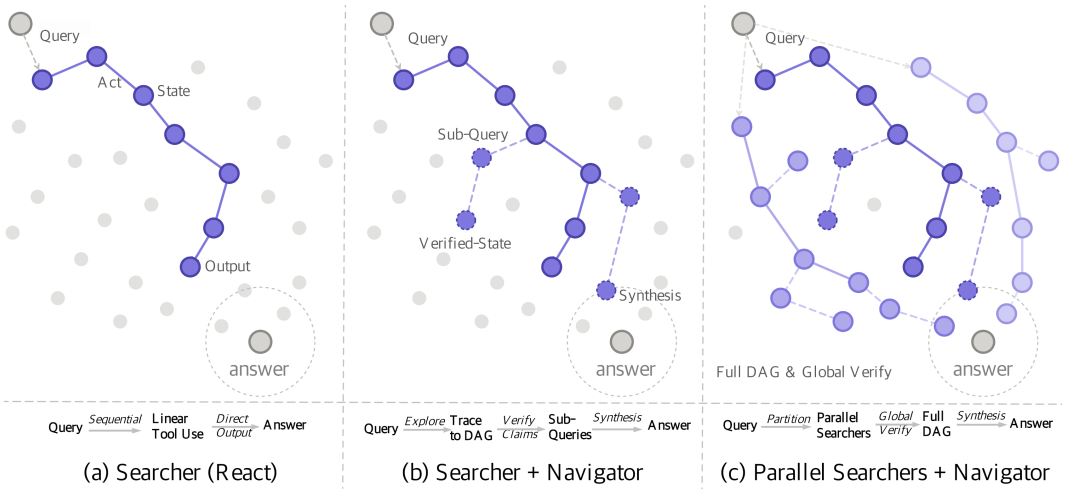

Figure 1: **Argus operating modes. (a)** A standalone Searcher executes one trajectory. **(b)** A Navigator assembles findings and requests targeted follow-up searches. **(c)** Multiple initial Searchers explore different query rewrites before graph-guided follow-up and synthesis.

The two components are trained separately. The Searcher remains a standard tool-using agent and is held fixed while the Navigator is optimized through supervised warm-starting and reinforcement learning. We evaluate the same trained Navigator with one or multiple initial Searchers and with different Searcher backbones. Either mode can request follow-up searches, so the initial Searcher count is only one part of its inference budget.

Across eight answer-level benchmarks, Argus improves over the standalone 35B-A3B Searcher by 5.5 percentage points on average with one initial Searcher and by 12.7 points with eight. Increasing initial parallelism to 64 yields 86.2% on BrowseComp. At this configuration, the final synthesis graph view contains approximately 21.5K tokens, compared with 25.6M cumulative Searcher tokens. Controlled changes to the synthesis representation further show a 5.2-point benefit of the full graph over text-only input with the same Searcher rollouts. Additional 200-question BrowseComp diagnostics compare prompt-only, SFT-only, and fully trained Navigators, and targeted follow-ups against no follow-up or independent exploration. These experiments test the complete pipeline, its synthesis representation, and its training and follow-up components.

2 Argus: Agentic Evidence Assembly

Argus consists of a *Searcher* and a *Navigator* connected through a shared directed evidence graph $\mathcal{G}$. The Searcher collects information through independent tool-using rollouts; the Navigator extracts evidence and tentative claims, requests further searches, and synthesizes an answer. The output is an answer a and the graph used to produce it. Figure 2 summarizes the three stages.

(I) Searching. The Navigator rewrites the input question q into K queries that emphasize different angles of inquiry. Solo mode uses $K = 1$ initial Searcher; parallel mode uses multiple initial Searchers. Each Searcher returns its trajectory to the Navigator.

(II) Verifying and assembling. The Navigator extracts evidence and claims from returned trajectories and relates them through support and contradiction edges. It inspects the graph for under-supported claims, conflicting findings, and unanswered parts of q, then dispatches targeted follow-up queries. This process continues until the Navigator stops or the available budget is exhausted.

(III) Synthesizing. The Navigator clears the working context from the preceding loop and generates its final answer from $(q, \mathcal{G})$. The graph retains the evidence and source links used in synthesis.

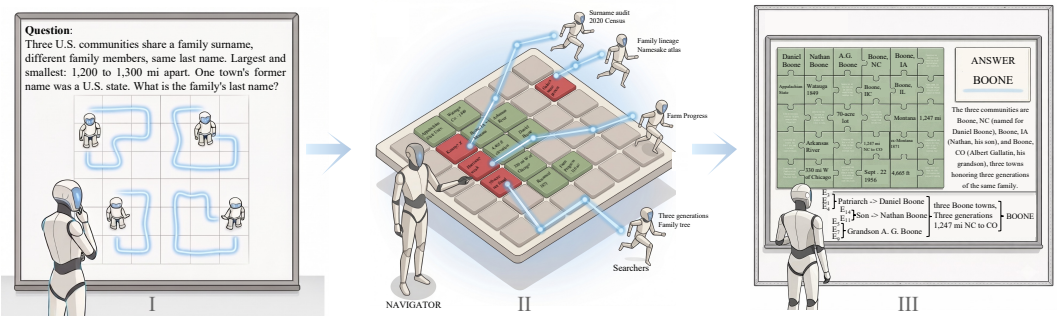

Figure 2: **Argus on a BrowseComp-style question. (I) Exploration**: Searchers execute independent ReAct rollouts. **(II) Verification**: the Navigator organizes retrieved evidence and tentative claims (green: corroborated findings; red: discarded probes) and dispatches targeted searches. **(III) Synthesis**: the Navigator uses the assembled graph and its source-linked evidence E_i to produce an answer.

2.1 SEARCHER

A Searcher takes one query and returns a trajectory H. It follows ReAct (Yao et al., 2023b): at step t, it generates a thought τ_t, takes an action $\alpha_t = (\alpha_t^m, \alpha_t^p)$, and receives an observation o_t. The action vocabulary is $\alpha_t^m \in \{\text{SEARCH}, \text{VISIT}, \text{ANSWER}\}$. SEARCH returns the top ten web results; VISIT returns an extractive page summary; and ANSWER terminates the rollout with an answer and a rationale. The Navigator receives the complete trajectory $H = (\tau_0, \alpha_0, o_0, \ldots, \tau_T, \alpha_T, o_T)$. A Searcher maintains its own within-rollout history but carries no state across invocations. It does not read $\mathcal{G}$ or communicate directly with other Searchers, so queries in the same batch can run concurrently.

2.2 NAVIGATOR

Observing trajectories and growing the graph. The Navigator maintains

$$\mathcal{G} = (E, C, \mathcal{A}), \qquad \mathcal{A} \subseteq (E \cup C) \times C \times \{+1, -1\}, \tag{1}$$

where E contains retrieved findings with source URLs and C contains tentative claims, including intermediate conclusions and Searcher answers. An arc attaches an evidence or claim node to a claim with a support $(+1)$ or contradict (-1) label. The intended representation is acyclic; the set of permitted arc types in Eq. 1 alone does not enforce acyclicity.

Unlike trees over one agent's reasoning steps (Yao et al., 2023a) or entity graphs over a static corpus (Edge et al., 2024; Gutiérrez et al., 2024), this graph consolidates information from independent Searcher trajectories. The Navigator parses each return and adds evidence, claims, and their relations. Source-URL deduplication reduces repeated records of the same page; distinct URLs can still share a source.

Each claim receives a policy-assigned status: *supported*, *contradicted*, or *unverified*. The Navigator assigns these labels from the incoming evidence, considering corroboration, conflicts, and source diversity. There is no fixed support-count threshold. Both the relations and the statuses are model predictions and may be incorrect.

Verifying claims and dispatching searches. The Navigator examines the graph jointly and produces a batch of queries $\mathcal{V} = \{v_1, \ldots, v_m\}$. Queries can seek independent corroboration for an under-supported claim, evidence resolving a contradiction, or information about an unanswered part of q. Each query is assigned to a Searcher; returned trajectories are incorporated before the next decision. The Navigator thus combines question decomposition with evidence-seeking verification, related to verification and follow-up questioning in Dhuliawala et al. (2024); Press et al. (2023).

The Navigator stops by emitting an end-of-loop token or reaching the compute cap. We distinguish the initial Searcher count K from the total dispatch count $D = K + \sum_r m_r$, where m_r is the number

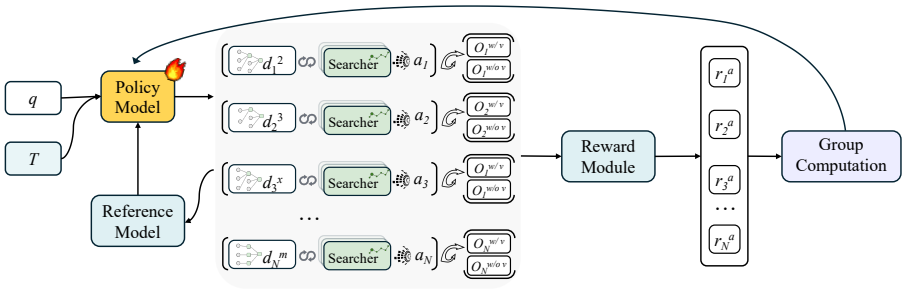

Figure 3: **Argus training pipeline.** For a question q and a pre-collected Searcher trajectory T, the Navigator samples N rollouts. Each produces a post-verification synthesis $y^{\star}_{\mathrm{w/v}}$ and a shadow synthesis $y^{\star}_{\mathrm{w/o\,v}}$ from the pre-verification graph. Judge scores determine the clipped reward in Eq. 3. Only Navigator-generated tokens receive policy gradients; the shadow pass is excluded from the training sequence.

of follow-up queries in round r. Under a total dispatch cap B, both initial and follow-up searches count toward $D \leq B$. Consequently, Solo denotes one initial Searcher, not necessarily one total rollout, and equal K does not imply equal total inference compute.

Synthesizing over the graph. After the loop, the Navigator generates

$$y^{\star} = \pi_{\mathrm{syn}}(q, \mathcal{G}), \tag{2}$$

where π_{syn} is its synthesis policy. This pass uses a compact graph view rather than the complete accumulated trajectories. The view groups evidence by source and includes claim status and corroboration annotations to guide synthesis. Its input size depends on the serialized graph, which grows as findings and claims are added. The final pass reads this graph instead of the raw rollout history; earlier passes still process trajectories and verification returns.

3 Navigator Training

Trained components. Searcher and Navigator are separately trained components initialized from Qwen3.5-35B-A3B (Qwen Team, 2026) (35B total, 3B active parameters). The Searcher receives supervised fine-tuning on trajectories produced through the WebSailor-v2 (Li et al., 2025b) pipeline and remains fixed during Navigator training. The Navigator is warm-started on graph-extraction and synthesis traces, then optimized with a group-relative, PPO-clipped policy objective (Shao et al., 2024). The training setup and hyperparameters are given in Appendix A.

Paired synthesis and reward. For each rollout, synthesis is run with the same Navigator weights on two graph states: $y^{\star}_{\mathrm{w/v}} = \pi_{\mathrm{syn}}(q, \mathcal{G}_{\mathrm{post}})$ and $y^{\star}_{\mathrm{w/o\,v}} = \pi_{\mathrm{syn}}(q, \mathcal{G}_{\mathrm{pre}})$. The post-verification graph includes the returns from additional searches; the pre-verification graph is the state before those searches. The shadow synthesis is used for scoring and carries no gradient. With training-judge scores $R_{\mathrm{w/v}}$ and $R_{\mathrm{w/o\,v}}$, the reward is

$$R_i = \mathrm{clip}\big(R_{\mathrm{w/v}} + \lambda(R_{\mathrm{w/v}} - R_{\mathrm{w/o\,v}}), 0, 1\big), \qquad \lambda = 0.5. \tag{3}$$

For binary judge scores, clipping makes this reward exactly $R_{\mathrm{w/v}}$ for any $\lambda \geq 0$ (Appendix A). In that setting, the objective rewards the final answer, with no additional contrastive bonus. We therefore make no claim of a separate learning benefit from λ.

Group-relative optimization. For each (q, T), the behavior policy $\pi_{\theta_{\mathrm{old}}}$ generates N Navigator rollouts with rewards $\{R_i\}_{i=1}^{N}$. The group-centered advantage is

$$A_i = R_i - \frac{1}{N}\sum_{j=1}^{N} R_j. \tag{4}$$

This mean-centered form does not include division by the group's reward standard deviation. Policy optimization uses PPO clipping with threshold ϵ and a KL penalty with coefficient β to a fixed reference policy. Only the Navigator's generated actions and synthesis tokens contribute to the policy loss; Searcher trajectories, tool observations, and shadow-synthesis tokens do not.

Rollout structure and parallel inference. Training uses one pre-collected Searcher trajectory T per question. The observation stage processes it through sliding windows and incrementally builds $\mathcal{G}_{\text{pre}}$. Verification queries are then sent to the fixed Searcher, and their returns update the graph to $\mathcal{G}_{\text{post}}$ before final synthesis. At inference time, the Navigator can assemble the graph from multiple initial trajectories. Section 4.3 evaluates this transfer to parallel search.

4 Experiments

4.1 Experimental Setup

Benchmarks. We evaluate answer-level information seeking on eight benchmarks. **BrowseComp** (Wei et al., 2025) and its Chinese counterpart **BrowseComp-ZH** (Zhou et al., 2025) probe multi-step web browsing on adversarially constructed factual questions that resist single-hop search. **xbench DeepSearch-2510** (Chen et al., 2025) targets deep search and tool use through professionally annotated Chinese tasks with dynamic updates. **GAIA** (Mialon et al., 2023) stresses general assistant capabilities that combine tool use, multi-hop reasoning, and web search across real-world question types. **SEAL-0** (Pham et al., 2026) is the main challenge track of SealQA, testing factual reasoning with noisy, conflicting, or unhelpful search results. **Humanity's Last Exam** (Phan et al., 2025) probes expert-level knowledge across science, mathematics, law, and medicine. **FrontierScience-Olympiad** (Wang et al., 2026) targets Olympiad-difficulty problems in physics, chemistry, and biology, written and verified by competition-level experts, while **FrontierScience Research** (Wang et al., 2026) extends this to PhD-level research sub-problems. This suite covers search and expert reasoning, but does not evaluate long-form research-report generation.

Compared systems. Table 1 includes proprietary agents (OpenAI, 2025a; Anthropic, 2026; Google DeepMind, 2026; Bytedance Seed, 2026), open-weight agents (GLM-5 Team et al., 2026; Moonshot AI, 2026; Qwen Team, 2026; DeepSeek-AI et al., 2026), and specialized research agents (Tongyi DeepResearch Team et al., 2025; MiroMind Team et al., 2026; Zeng et al., 2026). Unmarked external results come from their reports; † denotes our reproductions using `search` and `visit`. External results differ in tools, subsets, and inference budgets. For example, Seed reports text-only HLE, whereas Gemini and Claude report the full multimodal set. Appendix B describes the comparison scope.

Argus configurations. Both components use Qwen3.5-35B-A3B (Qwen Team, 2026), with the Searcher fine-tuned via WebSailor-v2 (Li et al., 2025b). *Searcher* is the standalone ReAct agent. *Argus-Solo* adds the Navigator to one initial Searcher; *Argus-Parallel* uses $K{=}8$ initial Searchers. Both may dispatch further Searchers during verification: K counts initial rollouts, not total compute. Our evaluations report mean accuracy over three inference seeds; externally sourced scores follow their original reports. Table 2 reports a separate 200-question diagnostic evaluation with per-seed scores in Appendix C; these results are not pooled with the benchmark tables.

4.2 Main Results

Consistent gains over the same Searcher. Argus-Solo and Argus-Parallel improve over the standalone Searcher on all eight benchmarks. Averaging the benchmark scores with equal weight gives 59.7% for Solo and 66.9% for Parallel, versus 54.2% for Searcher: gains of 5.5 and 12.7 percentage points. Parallel adds 7.2 points over Solo. The component comparisons below examine the contributions of Navigator training and follow-up search.

Comparison with reported systems. Argus-Parallel has the highest listed score on BrowseComp-ZH, GAIA, Seal-0, xbench DeepSearch-2510, and FrontierScience Olympiad. At $K{=}64$, it reaches 86.2% on BrowseComp (Section 4.3), numerically close to Gemini-3.1-Pro's reported 85.9% under a different evaluation setup.

Table 1: Answer accuracy (%) on eight benchmarks. †: reproduced with `search`/`visit`, without context management. ‡: prior work uses a 100-question subset; Argus uses the full set. Other external scores are from official reports. External budgets and tool settings are not matched; bold denotes the largest listed value, not a controlled ranking.

System	Browse Comp	Browse Comp-ZH	GAIA	Seal-0	xbench 2510	HLE	FrontierSci. Olympiad	FrontierSci. Research
Proprietary Agents								
GPT-5.2	65.8	—	—	—	—	45.5	77.1	25.2
Claude-4.6-Opus	83.7	66.8†	75.0†	50.0†	—	53.0	73.0†	23.3†
Gemini-3.1-Pro	**85.9**	74.0†	80.6†	42.5†	—	51.4	76.7†	20.0†
Seed-2.0-Pro	77.3	82.4	79.6†	49.5	—	**54.2**	74.0	**33.4**†
Open-Weight Agents								
GLM-5.0	75.9	72.7	70.0†	33.3†	—	50.4	62.0†	8.3†
Kimi-K2.6	83.2	—	78.6†	42.0†	—	54.0	—	—
Qwen3.5-35B-A3B	42.1†	47.8†	80.0†	43.2†	—	39.5†	68.0†	3.3†
Qwen3.5-397B-A17B	78.6	70.3	—	46.9	—	48.3	60.6	11.7†
DeepSeek-V4-Pro-Max	83.4	—	65.3†	—	—	48.2	75.0†	—
Specialized Deep Research Agents								
Tongyi-DeepResearch	43.4	46.7	70.9	—	55.0	32.9	—	—
MiroThinker-1.7	74.0	75.3	82.7	53.0	62.0	42.9	71.5	8.8†
Parallel Agents								
Tongyi-DeepResearch Heavy	69.0‡	55.0	72.8	—	—	—	—	—
GLM-4.5 Heavy	54.0‡	49.0	66.0	—	—	—	—	—
Our Searcher and Argus Configurations								
Searcher-35B-A3B	55.0	62.3	84.5	48.9	62.0	43.2	72.0	5.4
Argus-35B-A3B (Solo)	62.2	74.4	88.0	53.2	67.0	44.2	75.0	13.2
Argus-35B-A3B (Parallel)	74.5	**83.4**	**93.2**	**56.2**	**73.0**	49.8	**80.0**	25.0

Navigator training and follow-up search. Table 2 compares five configurations on BrowseComp, with 200 questions per inference run. Prompt-only uses the Navigator protocol without SFT or RL; SFT-only omits RL. No follow-up verification disables verification-driven Searcher dispatches. Extra independent search replaces targeted follow-ups with independent exploration. Full Argus scores 74.50%, exceeding prompt-only by 7.17 points, SFT-only by 4.83, and no follow-up verification by 5.17. It also exceeds extra independent search by 4.17 points.

Table 2: Navigator ablations on BrowseComp: 200 questions per run, three inference seeds. Accuracy is mean $\pm$ sample SD; Δ is the gap to Full Argus, in percentage points. Per-seed scores are in Appendix C.

	Configuration	Accuracy (%)	Δ
A1	Full Navigator (SFT + RL)	$\mathbf{74.50 \pm 2.00}$	—
A2	Prompt-only Navigator	67.33 ± 1.15	7.17
A3	SFT-only Navigator	69.67 ± 1.53	4.83
A4	No follow-up verification	69.33 ± 1.26	5.17
A5	Extra independent search	70.33 ± 2.08	4.17

Both training stages and targeted follow-up search improve the mean score in these runs. The ordering is consistent with a benefit from evidence-directed queries, but does not isolate query selection from differences in search effort. Realized token use and latency were not measured. Standard deviations describe variation across inference seeds; they do not include retraining variability.

4.3 Analysis

Scaling with Searcher budget. Figure 4 plots BrowseComp accuracy over a 64× range of cumulative Searcher tokens (0.4M–25.6M). The 55.0% endpoint is the standalone Searcher reference; Solo scores 62.2% in Table 1. The largest Argus configuration reaches 86.2%. We show the measured points without fitting or extrapolating a scaling law.

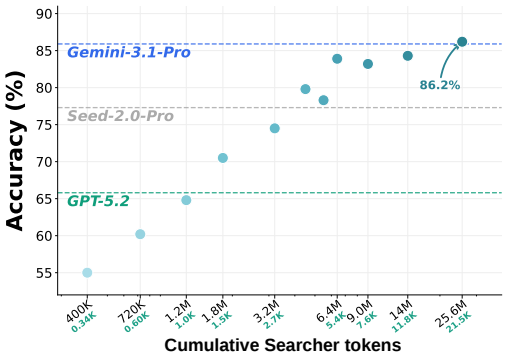

Figure 4: BrowseComp scaling. The horizontal axis counts Searcher tokens; green labels give final graph context. Dashed lines are external reference scores.

At this scale, final synthesis uses a graph view of approximately 21.5K tokens, within the Navigator's 128K context window. The ratio of cumulative Searcher tokens to final graph tokens is about 1,190:1. This ratio concerns the synthesis input: it excludes Navigator computation for reading trajectories, constructing the graph, and requesting verification. The graph can grow as further evidence is added.

Transfer across Searcher backbones. Table 3 pairs the same Navigator with three Searcher backbones under five inference configurations. The Navigator is trained with Searcher-35B-A3B and reused with DeepSeek and Seed-2.0-Pro without retraining. Argus-Parallel achieves the highest score for each backbone, exceeding the next-best listed configuration by 12.3, 9.5, and 3.8 points, respectively.

Table 3: Argus with different Searcher backbones on BrowseComp (mean Pass@1, %). K specifies initial Searchers; equal K does not imply equal total inference budgets.

Searcher backbone	Searcher	Argus-Solo	Majority-Vote (K=8)	LLM-Aggregation (35B-K=8)	Argus-Parallel (K=8)
Searcher-35B-A3B	55.0	62.2	56.2	56.5	**74.5**
DeepSeek-V4-Flash-Max	64.0	68.0	60.0	69.0	**78.5**
Seed-2.0-Pro	70.2	78.6	67.0	73.8	**82.4**

Argus-Solo exceeds eight-rollout Majority Vote for each backbone, but trails LLM-Aggregation by 1.0 point on DeepSeek. Argus-Parallel exceeds LLM-Aggregation by 18.0, 9.5, and 8.6 points. These results show that the Navigator remains useful when the Searcher backbone changes. Follow-up dispatches make total budgets unequal; LLM-Aggregation is also distinct from the tool-using asymmetric verifier of Zeng et al. (2026).

Ablation on graph representation. Table 4 holds Searcher rollouts fixed and changes the input to final synthesis. Bare graph adds connectivity and coarse claim status to Text only, improving accuracy by 2.7 points. Full DAG adds typed relations and corroboration annotations for another 2.5 points, or 5.2 points over Text only. Thus, the representation matters even with the retrieved information held fixed. Because connectivity and status change together, the Bare graph comparison cannot attribute the gain to topology alone.

4.4 Limitations

The largest configuration uses 25.6M Searcher tokens per question, and we have not measured an end-to-end cost or latency advantage. Component ablations cover one benchmark and three inference seeds. A matched-budget tool-using verifier remains an important missing comparison. We have also not measured graph-edge accuracy or how often follow-ups correct answers, introduce errors, or resolve contradictions. Extraction mistakes and dependent sources can produce false corroboration despite explicit source links. Finally, our evaluations cover answer-level information seeking, not long-form research reports.

Table 4: Synthesis-input ablation (K=8, BrowseComp). All variants share identical Searcher rollouts; only the representation presented to final synthesis changes.

Variant	What the Navigator sees	BrowseComp
Full DAG	Evidence and claim nodes, typed support/contradict edges, claim status, and corroboration strength	**74.5**
Bare graph	Evidence and claim nodes, untyped connectivity, and coarse claim status; no corroboration-strength annotations	72.0
Text only	Evidence and claim text concatenated in extraction order; no edges or status	69.3

5 Related Work

Deep research and context management. Search-enabled agents combine retrieval and reasoning (OpenAI, 2025b; Citron, 2024; Li et al., 2025d). WebWalker uses an explore–critic framework (Wu et al., 2025b); WebThinker integrates a Deep Web Explorer (Li et al., 2025e); MiroThinker studies research-agent verification (MiroMind Team et al., 2026). Other work develops training data and information-seeking policies (Wu et al., 2025a; Li et al., 2025c;b; Liu et al., 2025; Tongyi DeepResearch Team et al., 2025). WebResearcher, AgentFold, and ReSum address long-horizon execution and context management (Qiao et al., 2025; Ye et al., 2025; Wu et al., 2025c). Argus maintains shared claim/evidence state across independent Searcher returns.

Parallel scaling and verification. Repeated sampling, self-consistency, selection, and search scale inference computation (Brown et al., 2024; Wang et al., 2023; Snell et al., 2024; Lightman et al., 2023; Hao et al., 2023). ParallelMuse applies parallel thinking to information seeking (Li et al., 2025a); aggregation agents inspect candidates and search across rollouts (Lee et al., 2026). Zeng et al. (2026) score candidates with a tool-using verifier. Argus links claim assessments to subsequent searches through a persistent graph. Our voting and LLM-aggregation baselines do not reproduce that verifier under matched total compute.

Multi-agent coordination. AutoGen, CAMEL, and MetaGPT coordinate agents (Wu et al., 2023; Li et al., 2023; Hong et al., 2024); debate and societies-of-mind methods combine perspectives through communication (Du et al., 2024; Zhuge et al., 2023). Kimi K2.5 uses agent swarms for parallel delegation (Kimi Team et al., 2026). Argus coordinates searches around evidence coverage: independent Searchers return to a Navigator that maintains shared state and selects follow-up queries.

Structured reasoning and factuality. Tree and Graph of Thoughts structure reasoning (Yao et al., 2023a; Besta et al., 2024); GraphRAG and HippoRAG structure retrieval (Edge et al., 2024; Gutiérrez et al., 2024). Self-Refine, self-ask, and Chain-of-Verification study feedback, decomposition, and verification (Madaan et al., 2023; Press et al., 2023; Dhuliawala et al., 2024). FActScore, SAFE, and FacTool assess factual claims using external evidence or tools (Min et al., 2023; Wei et al., 2024; Chern et al., 2023). In Argus, source-linked claims and support/contradiction relations guide further search as well as final synthesis.

6 Conclusion

Argus connects independent Searcher rollouts through a shared evidence graph, allowing a Navigator to request missing evidence and synthesize source-linked answers. Across eight benchmarks, the complete pipeline improves on the standalone Searcher by 5.5 points with one initial Searcher and 12.7 with eight. At 64 initial Searchers it reaches 86.2% on BrowseComp, with a 21.5K-token final graph view representing 25.6M cumulative Searcher tokens. The component ablations favor Navigator training and targeted follow-up search. The reliability of individual graph relations and performance under matched total inference budgets remain open.

AI USE STATEMENT

The research system uses language models as Searchers, a Navigator, and a training-reward judge, as described in Sections 2–3 and Appendix A. Language models also generate the synthetic Searcher training trajectories. During manuscript revision, generative AI tools assisted with language editing, bibliography checking against primary sources, experimental-design suggestions, interpretation of supplied results, consistency and mathematical checks, and LaTeX formatting. AI-generated expectations and illustrative numbers are excluded from the reported measurements. The authors remain responsible for the manuscript, its references, and its empirical claims.

ETHICS STATEMENT

Web-based research agents can retrieve inaccurate, biased, private, unsafe, or copyrighted material. Source-linked evidence makes selected reasoning inputs inspectable but does not ensure source reliability, factual correctness, permission to reuse content, or safe downstream decisions. Deployment requires appropriate data handling, access controls, and human oversight, particularly in high-stakes settings. The substantial inference budgets used in large parallel configurations also warrant consideration of resource costs.

REPRODUCIBILITY STATEMENT

Sections 2 and 3 describe the component interfaces, graph representation, and training objective. Appendix A records the available training settings and the interpretation of the clipped reward; Appendix B clarifies dispatch accounting and comparison scope, and Appendix C lists each diagnostic seed score. Section 4 distinguishes controlled representation comparisons from external reported scores and configurations with unmatched total inference budgets. Appendix D gives an abbreviated qualitative trace. These descriptions do not replace per-example evaluation outputs or a complete executable release.

REFERENCES

Anthropic. System card: Claude Opus 4.6, 2026. URL https://www-cdn.anthropic.com/6a5fa276ac68b9aeb0c8b6af5fa36326e0e166dd.pdf.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michał Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, and Torsten Hoefler. Graph of thoughts: Solving elaborate problems with large language models. In *AAAI Conference on Artificial Intelligence*, 2024. URL https://ojs.aaai.org/index.php/AAAI/article/view/29720.

Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V Le, Christopher Ré, and Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling. *arXiv preprint arXiv:2407.21787*, 2024. URL https://arxiv.org/abs/2407.21787.

Bytedance Seed. Seed2.0 model card: Towards intelligence frontier for real-world complexity, 2026. URL https://lf3-static.bytednsdoc.com/obj/eden-cn/lapzild-tss/ljhwZthlaukjlkulzlp/seed2/0214/Seed2.0%20Model%20Card.pdf.

Kaiyuan Chen, Yixin Ren, Yang Liu, Xiaobo Hu, Haotong Tian, Tianbao Xie, Fangfu Liu, Haoye Zhang, Hongzhang Liu, Yuan Gong, et al. xbench: Tracking agents productivity scaling with profession-aligned real-world evaluations. *arXiv preprint arXiv:2506.13651*, 2025. URL https://arxiv.org/abs/2506.13651.

I-Chun Chern, Steffi Chern, Shiqi Chen, Weizhe Yuan, Kehua Feng, Chunting Zhou, Junxian He, Graham Neubig, and Pengfei Liu. FacTool: Factuality detection in generative AI – a tool augmented framework for multi-task and multi-domain scenarios. *arXiv preprint arXiv:2307.13528*, 2023. URL https://arxiv.org/abs/2307.13528.

Dave Citron. Try Deep Research and our new experimental model in Gemini, your AI assistant. Google Blog, 2024. URL https://blog.google/products-and-platforms/products/gemini/google-gemini-deep-research/.

DeepSeek-AI et al. DeepSeek-V4: Towards highly efficient million-token context intelligence, 2026. URL https://arxiv.org/abs/2606.19348.

Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason Weston. Chain-of-Verification reduces hallucination in large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 3563–3578, 2024. URL https://aclanthology.org/2024.findings-acl.212/.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. In *International Conference on Machine Learning (ICML)*, 2024. URL https://proceedings.mlr.press/v235/du24e.html.

Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph RAG approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024. URL https://arxiv.org/abs/2404.16130.

GLM-5 Team, Aohan Zeng, Xin Lv, Zhenyu Hou, Zhengxiao Du, Qinkai Zheng, Bin Chen, Da Yin, Chendi Ge, Chenghua Huang, Chengxing Xie, et al. GLM-5: From vibe coding to agentic engineering. *arXiv preprint arXiv:2602.15763*, 2026. URL https://arxiv.org/abs/2602.15763.

Google DeepMind. Gemini 3.1 Pro: Model card, February 2026. URL https://deepmind.google/models/model-cards/gemini-3-1-pro/.

Bernal Jiménez Gutiérrez, Yiheng Shu, Yu Gu, Michihiro Yasunaga, and Yu Su. HippoRAG: Neurobiologically inspired long-term memory for large language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/6ddc001d07ca4f319af96a3024f6dbd1-Abstract.html.

Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang, and Zhiting Hu. Reasoning with language model is planning with world model. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023. URL https://aclanthology.org/2023.emnlp-main.507/.

Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. MetaGPT: Meta programming for a multi-agent collaborative framework. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/6507b115562bb0a305f1958ccc87355a-Abstract-Conference.html.

Kimi Team, Tongtong Bai, Yifan Bai, Yiping Bao, S. H. Cai, Yuan Cao, Ziwei Chai, Y. Charles, H. S. Che, Cheng Chen, Guanduo Chen, et al. Kimi K2.5: Visual agentic intelligence. *arXiv preprint arXiv:2602.02276*, 2026. URL https://arxiv.org/abs/2602.02276.

Yoonsang Lee, Howard Yen, Xi Ye, and Danqi Chen. Agentic aggregation for parallel scaling of long-horizon agentic tasks. *arXiv preprint arXiv:2604.11753*, 2026. URL https://arxiv.org/abs/2604.11753.

Baixuan Li, Dingchu Zhang, Jialong Wu, Wenbiao Yin, Zhengwei Tao, Yida Zhao, Liwen Zhang, Haiyang Shen, Runnan Fang, Pengjun Xie, Jingren Zhou, and Yong Jiang. ParallelMuse: Agentic parallel thinking for deep information seeking. *arXiv preprint arXiv:2510.24698*, 2025a. URL https://arxiv.org/abs/2510.24698.

Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. CAMEL: Communicative agents for "mind" exploration of large language model society. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. URL https://proceedings.neurips.cc/paper/2023/hash/a3621ee907def47c1b952ade25c67698-Abstract-Conference.html.

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. *arXiv preprint arXiv:2509.13305*, 2025b. 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 super-human reasoning for web agent. *arXiv preprint arXiv:2507.02592*, 2025c. URL https://arxiv.org/abs/2507.02592.

Xiaoxi Li, Guanting Dong, Jiajie Jin, Yuyao Zhang, Yujia Zhou, Yutao Zhu, Peitian Zhang, and Zhicheng Dou. Search-o1: Agentic search-enhanced large reasoning models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 5420–5438, 2025d. URL https://aclanthology.org/2025.emnlp-main.276/.

Xiaoxi Li, Jiajie Jin, Guanting Dong, Hongjin Qian, Yongkang Wu, Ji-Rong Wen, Yutao Zhu, and Zhicheng Dou. WebThinker: Empowering large reasoning models with deep research capability. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025e. URL https://papers.nips.cc/paper_files/paper/2025/hash/ae03bdef276132fae089692445725635-Abstract-Conference.html.

Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. *arXiv preprint arXiv:2305.20050*, 2023. URL https://arxiv.org/abs/2305.20050.

Junteng Liu, Yunji Li, Chi Zhang, Jingyang Li, Aili Chen, Ke Ji, Weiyu Cheng, Zijia Wu, Chengyu Du, Qidi Xu, et al. WebExplorer: Explore and evolve for training long-horizon web agents. *arXiv preprint arXiv:2509.06501*, 2025. URL https://arxiv.org/abs/2509.06501.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-Refine: Iterative refinement with self-feedback. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/91edff07232fb1b55a505a9e9f6c0ff3-Abstract-Conference.html.

Grégoire Mialon, Clémentine Fourrier, Craig Swift, Thomas Wolf, Yann LeCun, and Thomas Scialom. GAIA: A benchmark for general AI assistants. *arXiv preprint arXiv:2311.12983*, 2023. URL https://arxiv.org/abs/2311.12983.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. FActScore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 12076–12100, 2023. URL https://aclanthology.org/2023.emnlp-main.741/.

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

Moonshot AI. Kimi K2.6: Advancing open-source coding, 2026. URL https://www.kimi.com/blog/kimi-k2-6.

OpenAI. Update to GPT-5 system card: GPT-5.2, 2025a. URL https://openai.com/index/gpt-5-system-card-update-gpt-5-2/.

OpenAI. Introducing Deep Research. https://openai.com/index/introducing-deep-research/, 2025b.

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. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/a0e0f9c39b7f6d07a266a3326216ec40-Abstract-Conference.html.

Long Phan, Alice Gatti, Ziwen Han, et al. Humanity's last exam. *arXiv preprint arXiv:2501.14249*, 2025. URL https://arxiv.org/abs/2501.14249.

Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt, Noah A. Smith, and Mike Lewis. Measuring and narrowing the compositionality gap in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 5687–5711, 2023. URL https://aclanthology.org/2023.findings-emnlp.378/.

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. *arXiv preprint arXiv:2509.13309*, 2025. URL https://arxiv.org/abs/2509.13309.

Qwen Team. Qwen3.5: Accelerating productivity with native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024. URL https://arxiv.org/abs/2402.03300.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling model parameters. *arXiv preprint arXiv:2408.03314*, 2024. URL https://arxiv.org/abs/2408.03314.

Tongyi DeepResearch Team, Baixuan Li, Bo Zhang, Dingchu Zhang, Fei Huang, Guangyu Li, Guoxin Chen, Huifeng Yin, Jialong Wu, Jingren Zhou, Kuan Li, Liangcai Su, Litu Ou, Liwen Zhang, Pengjun Xie, Rui Ye, Wenbiao Yin, Xinmiao Yu, Xinyu Wang, Xixi Wu, Xuanzhong Chen, Yida Zhao, Zhen Zhang, Zhengwei Tao, Zhongwang Zhang, Zile Qiao, et al. Tongyi DeepResearch technical report. *arXiv preprint arXiv:2510.24701*, 2025. URL https://arxiv.org/abs/2510.24701.

Miles Wang, Robi Lin, Kat Hu, Joy Jiao, Neil Chowdhury, Ethan Chang, and Tejal Patwardhan. FrontierScience: Evaluating AI's ability to perform expert-level scientific tasks. *arXiv preprint arXiv:2601.21165*, 2026. URL https://arxiv.org/abs/2601.21165.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=1PL1NIMMrw.

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. *arXiv preprint arXiv:2504.12516*, 2025. URL https://arxiv.org/abs/2504.12516.

Jerry Wei, Chengrun Yang, Xinying Song, Yifeng Lu, Nathan Hu, Jie Huang, Dustin Tran, Daiyi Peng, Ruibo Liu, Da Huang, Cosmo Du, and Quoc V. Le. Long-form factuality in large language models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/937ae0e83eb08d2cb8627fe1def8c751-Abstract-Conference.html.

Jialong Wu, Baixuan Li, Runnan Fang, Wenbiao Yin, Liwen Zhang, Zhenglin Wang, Zhengwei Tao, Ding-Chu Zhang, Zekun Xi, Robert Tang, Yong Jiang, Pengjun Xie, Fei Huang, and Jingren Zhou. WebDancer: Towards autonomous information seeking agency. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=quJdphBcdP.

Jialong Wu, Wenbiao Yin, Yong Jiang, Zhenglin Wang, Zekun Xi, Runnan Fang, Linhai Zhang, Yulan He, Deyu Zhou, Pengjun Xie, et al. WebWalker: Benchmarking LLMs in web traversal. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10290–10305, 2025b. URL https://aclanthology.org/2025.acl-long.508/.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W. White, Doug Burger, and Chi Wang. AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*, 2023. URL https://arxiv.org/abs/2308.08155.

Xixi Wu, Kuan Li, Yida Zhao, Liwen Zhang, Litu Ou, Huifeng Yin, Zhongwang Zhang, Xinmiao Yu, Dingchu Zhang, Yong Jiang, Pengjun Xie, Fei Huang, Minhao Cheng, Shuai Wang, Hong Cheng, and Jingren Zhou. ReSum: Unlocking long-horizon search intelligence via context summarization, 2025c. URL https://arxiv.org/abs/2509.13313.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023a. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/271db9922b8d1f4dd7aaef84ed5ac703-Abstract-Conference.html.

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 (ICLR)*, 2023b. URL https://arxiv.org/abs/2210.03629.

Rui Ye, Zhongwang Zhang, Kuan Li, Huifeng Yin, Zhengwei Tao, Yida Zhao, Liangcai Su, Liwen Zhang, Zile Qiao, Xinyu Wang, Pengjun Xie, Fei Huang, Siheng Chen, Jingren Zhou, and Yong Jiang. AgentFold: Long-horizon web agents with proactive context management, 2025. URL https://arxiv.org/abs/2510.24699.

Weihao Zeng, Keqing He, Chuqiao Kuang, Xiaoguang Li, and Junxian He. Pushing test-time scaling limits of deep search with asymmetric verification. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=hxL4Uf9tR3.

Peilin Zhou, Bruce Leon, Xiang Ying, Can Zhang, Yifan Shao, Qichen Ye, Dading Chong, Zhiling Jin, Chenxuan Xie, Meng Cao, Yuxin Gu, Sixin Hong, Jing Ren, Jian Chen, Chao Liu, and Yining Hua. BrowseComp-ZH: Benchmarking web browsing ability of large language models in chinese. *arXiv preprint arXiv:2504.19314*, 2025. URL https://arxiv.org/abs/2504.19314.

Mingchen Zhuge, Haozhe Liu, Francesco Faccio, Dylan R. Ashley, Róbert Csordás, Anand Gopalakrishnan, Abdullah Hamdi, Hasan Abed Al Kader Hammoud, Vincent Herrmann, Kazuki Irie, Louis Kirsch, Bing Li, Guohao Li, Shuming Liu, Jinjie Mai, Piotr Piękos, Aditya Ramesh, Imanol Schlag, Weimin Shi, Aleksandar Stanić, Wenyi Wang, Yuhui Wang, Mengmeng Xu, Deng-Ping Fan, Bernard Ghanem, and Jürgen Schmidhuber. Mindstorms in natural language-based societies of mind. *arXiv preprint arXiv:2305.17066*, 2023. URL https://arxiv.org/abs/2305.17066.

A Training Details

Searcher. The Searcher is initialized from Qwen3.5-35B-A3B (Qwen Team, 2026) and supervised-fine-tuned on approximately 10K trajectories synthesized through the WebSailor-v2 pipeline. No Argus-specific reinforcement learning is applied to this component. These 10K trajectories are Searcher training data, not the size of the Navigator SFT dataset.

Navigator initialization and SFT. The Navigator is initialized separately from the same base checkpoint and warm-started on graph-extraction and synthesis traces, using learning rate 10^{-5} and batch size 64. The checkpoint with the lowest held-out loss initializes RL.

Navigator RL data. RL uses 5,298 multi-hop information-seeking questions paired with verified answers and pre-collected Searcher trajectories. The filtering procedure removes training questions with named entities overlapping those in the eight evaluation benchmarks. Entity-level filtering is a contamination check, not a guarantee against paraphrased overlap or exposure during base-model pretraining. We do not use training-time diagnostic curves as evidence of test-set generalization.

Navigator rollouts. Graph observation uses sliding windows of 15 trajectory rounds, with at most eight windows, matching the window range used to construct SFT examples. Verification returns update the graph before synthesis. The shadow synthesis is computed separately for scoring and is excluded from the training sequence. Only Navigator-generated tokens carry policy gradients; pre-collected trajectories, verification returns, and other external inputs are masked. A state-machine check rejects rollouts that violate the graph output format before they proceed. Syntactic validation alone does not validate the evidence, edge semantics, or final answer.

Optimization and compute. The learning rate is 10^{-6}. Each rollout batch contains 64 prompts with $N = 8$ rollouts per prompt (512 samples); the oversampling batch is 128 and sampling temperature is 1.0. The clipping threshold is $\epsilon = 0.2$, the KL coefficient is $\beta = 0.001$, and the reward coefficient is $\lambda = 0.5$. The maximum response length is 135,168 tokens. Training runs for 100 rollout steps with sample and verification timeouts of 600 and 1,200 seconds, respectively. Training uses 64 H200 GPUs and takes approximately 1.5 days, including rollout generation and policy updates.

Interpreting the clipped reward. The training judge is DeepSeek-V3.1-Chat. The score support and clipping operation matter for interpreting Eq. 3. For binary scores and any $\lambda \geq 0$, the reward has the following values:

$R_{\mathrm{w/v}}$	$R_{\mathrm{w/o\,v}}$	R_i
0	0	0
0	1	0
1	0	1
1	1	1

Thus the clipped expression cannot assign an additional verification bonus with binary correctness scores. For non-binary scores, a difference can survive clipping, but its contribution depends on the score distribution and saturation at the clipping boundaries. The paired-score curves alone therefore do not establish a distinct learning benefit from λ. Furthermore, equal pre/post graph inputs do not force two independently sampled synthesis outputs to receive equal scores.

B Evaluation Setup

Dispatch budgets. K denotes initial Searcher invocations. The total dispatch count includes these and all verification-driven follow-ups: $D = K + \sum_r m_r$. A total-dispatch cap B bounds D, so configurations with $K = B$ leave no capacity for additional Searcher dispatches. A reported initial count such as $K = 64$ must not be interpreted as a measurement of total dispatches or of total system tokens. Searcher token consumption excludes Navigator processing unless explicitly stated otherwise.

Evaluation metrics and comparison scope. We report the benchmark scores used in Table 1. Benchmark-specific judging procedures are separate from the training-only reward judge. External baseline numbers are taken from their source reports; entries marked with a dagger are author reproductions using `search` and `visit`. Differences in tools, benchmark subsets, judges, evaluation dates, and inference budgets limit direct comparison with externally reported numbers. In particular, using a benchmark's reported metric does not by itself establish an identical evaluation configuration or a matched-compute comparison.

C Component Diagnostic Scores

Table 5 gives the individual inference-run scores underlying Table 2. Each run evaluates 200 questions. We compute the mean as $\bar{x} = \frac{1}{3}\sum_{s=1}^{3} x_s$ and sample standard deviation as $s_x = \sqrt{\frac{1}{2}\sum_{s=1}^{3}(x_s - \bar{x})^2}$, with both in percentage points. Inference repeats are not independent training replicates and must not be treated as 600 distinct evaluation questions.

Table 5: Individual accuracy scores (%) for the component diagnostics. Seed columns index inference runs.

Configuration	Questions/run	Seed 1	Seed 2	Seed 3
Full Navigator (SFT + RL)	200	74.5	72.5	76.5
Prompt-only Navigator	200	68.0	66.0	68.0
SFT-only Navigator	200	70.0	68.0	71.0
No follow-up verification	200	70.5	68.0	69.5
Extra independent search	200	72.0	71.0	68.0

The full configuration exceeds each alternative in all three reported columns. The table supplies descriptive evidence of the ordering, not a paired significance test: per-question predictions and covariance are not reported. It also does not establish equal realized compute across configurations. The independent-search comparison concerns query allocation rather than a separately trained, tool-using verifier.

D CASE STUDY

The following abbreviated case record shows selected Searcher returns and Navigator graph updates. The displayed entries summarize the record rather than reproduce a complete raw log; omitted rounds and nodes account for gaps in the identifiers. Claim nodes are written as C_i to match Eq. 1. This example illustrates the workflow, not a quantitative validation of graph accuracy.

Verification overturns an incorrect candidate

Question. What is the name of the individual who received a "special mention" in the acknowledgments section of a thesis submitted between 2018 and 2020 to a US university established in the 1800s? The author's then-graduate advisor was listed in the Providence Business News 40 Under Forty list. The author received the Saint Elmo Brady Award for Outstanding Achievement in Science from the URI Black Scholar Awards Committee. The individual contributed to a chapter of two books, both published in 2017.

Ground truth: Nicholas Constant.

Searcher trace H_1.

```
[thought] Five overlapping constraints. Start with the most
          distinctive: the Saint Elmo Brady Award at URI.
[search]  "URI Saint Elmo Brady Award recipient"
[obs]     ThankGod Ugochukwu (2018); Deborah Adenakan.
[search]  "URI Saint Elmo Brady Award PBN 40 Under Forty"
[obs]     Mentions Jesse Duroha alongside Gretchen Macht (PBN 2019).
[visit]   bipartisanpolicy.org/person/gretchen-macht/
[obs]     Confirms Macht received PBN 40 Under Forty in 2019.
          ... 4 rounds elided ...
[answer]  Jesse Duroha.
```

Navigator parses H_1 into $\mathcal{G}_1$:

Evidence: E_1 (Ugochukwu 2018), E_2 (Adenakan), E_3 (Roxbury PBN 2022), E_4 (Macht PBN 2019), E_5 (Duroha SEB).

Claim: C_1 = "Jesse Duroha is the answer." Status: `unverified` (corroborated by 2 sources, but 2018–2020 window and 2017 book chapters not touched).

Verify queries:

v_1: confirm Jesse Duroha satisfies all five constraints.

v_2: actual submission year of Duroha's thesis.

v_3 (alternative-hypothesis): are there other URI graduate students whose advisor is on PBN 40 Under Forty and who received the Saint Elmo Brady Award?

[verification batch dispatched, $|\mathcal{V}| = 3$; the trace responding to v_3 is shown below]

Searcher trace H_2 (responding to v_3).

```
[thought] Probe asks for OTHER URI students with the same award.
          Pivot away from Duroha; search the URI thesis archive.
[search]  "URI master thesis 2019 Saint Elmo Brady acknowledgments"
[obs]     Hit: Rassoul Diouf, "Data Analytics for Wearable IoT-Based
          Telemedicine", URI 2019.
[visit]   digitalcommons.uri.edu/.../article=2734&context=theses
[obs]     Summary: acknowledgments give Nicholas Constant a
          special mention. Advisor: Prof. Kunal Mankodiya.
          ... 3 rounds elided ...
[search]  "Kunal Mankodiya PBN 40 Under Forty"
[obs]     PBN 2017 honoree, director of URI Wearable Biosensing Lab.
[search]  "Nicholas Constant 2017 book chapter"
[obs]     Two 2017 chapters in Springer healthcare volumes.
[answer]  Nicholas Constant.
```

Navigator updates $\mathcal{G}_1 \to \mathcal{G}_2$:

New evidence: E_6 (Duroha thesis is 2023, contradicts window), E_9 (acknowledgments paragraph, decisive), E_{10}, E_{15} (Mankodiya PBN 2017), E_{12}, E_{13} (Diouf SEB 2017), E_{11}, E_{14} (Constant 2017 chapters).

New arcs to C_1: E_6, E_8, E_{16}–E_{19} all attach as contradict, changing the assigned status to `contradicted`.

New claim: $C_3 =$ "Nicholas Constant is the answer." Status: `supported`, with one recorded evidence node per constraint and no recorded contradictions.

Evidence trace over $\mathcal{G}_2$.

In the record, the thesis-related constraints identify the author and institution. At URI (whose predecessor was chartered in 1888), the 2017 Saint Elmo Brady awardee is Rassoul Diouf (E_{12}, E_{13}), and Diouf's master thesis was submitted in 2019, inside the 2018–2020 window (E_9). The same thesis (E_9) names Kunal Mankodiya as the graduate advisor, and Mankodiya is independently confirmed as the 2017 PBN 40 Under Forty honoree (E_{10}, E_{15}). The author and the thesis are thus linked by the recorded evidence.

The remaining constraints identify the answer. The acknowledgments section of Diouf's thesis, retrieved via a direct `visit` (E_9), is recorded as giving a special mention to Nicholas Constant. The record also associates Constant with two 2017 book chapters (E_{11}, E_{14}), completing the Navigator's constraint checks. These are the recorded assessments for this selected example, not independent annotations of the graph's correctness.

Final Answer: **Nicholas Constant**.

Source links. The displayed visits refer to Gretchen Macht's Bipartisan Policy Center profile and Rassoul Diouf's thesis record at the University of Rhode Island. The latter links to the thesis PDF identified by `article=2734&context=theses` in the abbreviated trace. URI's official history distinguishes its 1888 charter from its 1892 change to a college.