ORBIT: Scalable and Verifiable Data Generation for Search Agents on a Tight Budget

Nandan Thakur, Zijian Chen, Xueguang Ma & Jimmy Lin
David R. Cheriton School of Computer Science
University of Waterloo
{n3thakur,s42chen,x93ma,jimmylin}@uwaterloo.ca

Abstract

Search agents, which integrate language models (LMs) with web search, are becoming crucial for answering complex user queries. Constructing training datasets for deep research tasks, involving multi-step retrieval and reasoning, *remains challenging* due to expensive human annotation, or cumbersome pre-requisites. In this work, we introduce **ORBIT**, a training dataset with 20K reasoning-intensive queries with short verifiable answers, generated using a *frugal framework* without relying on paid API services. The modular framework relies on four stages: seed creation, question–answer pair generation, and two stages of verification: self and external. ORBIT spans 15 domains and each training pair requires 4–5 reasoning steps, with external search verification required from the complete web. We train Qwen3-4B as the base model on ORBIT using GRPO and evaluate it on Wikipedia question answering tasks. Extensive experiment results demonstrate that ORBIT-4B achieves strong performance among sub-4B LLMs as search agents, proving the utility of synthetic datasets. Our framework, code and datasets are open-sourced and available publicly.

castorini/orbit orbit-ai

1 Introduction

Large language models (LLMs) have been used for their reasoning capabilities beyond simple factoid queries in a new frontier we term as *deep search*. This setting requires interleaved reasoning, decomposing complex tasks, and reasoning across multiple sources of information, and it is now exposed as an endpoint in multiple commercial products (OpenAI, 2025; Gemini, 2025). Unlike traditional question-answering systems, answering *deep search* questions goes beyond a single-turn web search: it requires breaking down complex queries, conducting multiple retrieval steps, and iteratively planning searches and aggregating results from a search tool. Recent work training search agents, such as Search-R1 and InfoSeek (Jin et al., 2025; Xia et al., 2026) has shown that applying GRPO with verifiable rewards improves models' iterative reasoning and search tool use.

However, training data availability for *deep search* tasks is very limited. Search-R1 uses NQ (Kwiatkowski et al., 2019) and HotpotQA (Yang et al., 2018), however these contain simple retrieval queries that do not necessarily require multiple hops of searches for answering them. Collecting human annotations with verifiable answers for complex queries is also challenging: given a complex query, the uncertainty of web search makes the annotation burden impractically high. On the other hand, training datasets with complex search queries are either limited in training pairs, or require pre-requisites before generation, such as knowledge-graph construction, limiting their real-world applicability and usage (Wu et al., 2025; Sun et al., 2025b; Xi et al., 2025; Tao et al., 2026; Liu et al., 2025; Wolfson et al., 2026; Xia et al., 2026).

In this work, we introduce a frugal framework for constructing complex training data for search agents in a budget constrained manner, without relying on any paid API services.

Dataset Name	Source	Train Pairs	Complexity	Pre-requisite / Constraint
NQ (Kwiatkowski et al., 2019)	Wiki	300K+	Easy	None
HotpotQA (Yang et al., 2018)	Wiki	100K+	Easy	None
WebWalkerQA (Wu et al., 2025)	Web	14.3K	Easy	Web Traversal
SimpleDeepS. (Sun et al., 2025b)	Wiki & Web	871	Easy	Web Search & HTML Extract
InfoDeepSeek (Xi et al., 2025)	Web	245	Easy	Human Annotators
WebShaper (Tao et al., 2026)	Wiki	500	Hard	Wikipedia Hyperlinks
WebExplorer (Liu et al., 2025)	Web	100	Hard	Web Search & Obsfurcation
MoNaCo (Wolfson et al., 2026)	Wiki	1,315	Hard	Human Annotators
DeepDive (Lu et al., 2025)	KG	3,250	Hard	Knowledge Graph
InfoSeek (Xia et al., 2026)	Wiki & Web†	17.8K	Hard	Entity Linking & Obsfurcation
ORBIT (ours)	**Wiki & Web**	**20K**	**Hard**	**None**

Table 1: Comparison of ORBIT against other training datasets for search agents. Prior datasets require prior constraints such as a knowledge graph (KG), lack structural depth or remain limited in scale. ORBIT is a large-scale multi-hop dataset, requiring no prior constraints and containing complex queries, that can be scaled easily with a low budget.

The framework is modular and can be adopted easily for synthetic data generation. Existing synthetic deep search datasets commonly rely on graph structures and entity linkages, or are not fully open sourced, as shown in Table 1. Using our framework, we construct ORBIT (*Open-Web Reasoning for Information Retrieval Tasks*), a verified synthetic training dataset with narrative-style, i.e., long and reasoning-intensive questions with short verifiable answers, requiring multi-hops across the web.

Specifically, **ORBIT** is constructed by a four-stage framework that automatically constructs verifiable, multi-hop training pairs at scale: (1) *seed creation*: we expand 15 domains into large sets of Wikipedia categories and iterate linked pages, using page titles as seeds to ensure both head and long-tail coverage; (2) *question–answer pair generation*: conditioned on a seed, a search-enabled `DeepSeek-V3.1` composes an inverted question paired with a short verifiable answer; (3) *self-verification*: a search-enabled `DeepSeek-V3.1` assesses whether the proposed answer satisfies every atomic fact mentioned in the inverted question and provides supporting URLs as citations, discarding pairs with unverified answers or weakly supported subclaims; and (4) *external-verification*: two cascading LLM judges (`Qwen3-4B-Instruct-2507` and `gpt-oss-120b`) with scraped URL context, generate an answer and self-verify the generated answer as the judge with the ground-truth answer.

Using ORBIT, we fine-tune smaller and efficient LLMs (<4B parameters) as search agents using Qwen3-4B as the base model (Yang et al., 2025a), with GRPO (Shao et al., 2024). By mixing ORBIT with training datasets, ORBIT-4B demonstrates strong effectiveness to sub-4B search agent baselines on complex Wikipedia-based QA benchmarks, highlighting the potential of our framework for data generation to produce high-quality supervision for scaling small-sized search agents. We summarize our major contributions as follows:

1. We propose a frugal construction framework without requiring any pre-requisites or relying on expensive API's, to generate reasoning-intensive queries with short and verifiable answers on a tight budget.
2. We introduce ORBIT, one of the largest available datasets for training search agents with 20K+ verified question–answer pairs with a hard complexity, requiring verification both Wikipedia and Web domains for each question.
3. We empirically demonstrate the effectiveness of ORBIT through extensive experiments, showing that the proposed dataset yields significant performance gains on small-sized search agents. ORBIT-4B surpasses existing search agents with less than 4B model parameters by up to 9.0 EM accuracy on complex QA benchmarks on Wikipedia.
4. We open-source the framework, ORBIT dataset and the code for training search agents. Code is available on GitHub at: `https://github.com/castorini/orbit`, ORBIT dataset and search agents are available on Hugging Face at: `https://huggingface.co/orbit-ai`.

Question: What was the exact runtime (minutes) of the 2017 animated feature set inside a smartphone's messaging application, directed by a filmmaker previously known for sequels to popular children's franchises, featuring a protagonist whose facial expression malfunctions, with voice casting that includes a lead actor from a critically acclaimed sitcom, and produced by a studio that later won an Oscar for Spider-Man animation?
Answer: 86 minutes

Verification / Summary of Supporting Facts:
✓ **Animated Set:** *The Emoji Movie released in 2017, set inside a smartphone messaging app world called Textopolis.*
✓ **Filmmaker:** *Tony Leondis, sequels to franchises: Lilo & Stitch 2, Kung Fu Panda: Secrets of the Masters.*
✓ **Protagonist**: *Gene, the main character in The Emoji Movie, struggles with malfunctioning expressions.*
✓ **Voice Cast :** *T.J. Miller, who was a lead actor from the HBO sitcom titled Silicon Valley.*
✓ **Voice Casting Studio :** *Sony Pictures Animation, that won an Oscar for Spider-Man: Into the Spider-Verse.*

Table 2: An training data example showing the question, answer and verification (for the reader) from the *TV Shows & Movies* domain in the ORBIT dataset. Colored spans mark the distinct factual clues embedded in the question; the verification rows confirm each clue with key supporting evidence (blue). More dataset examples are provided in Appendix E.

2 Related Work

2.1 Reinforcement Learning and Search-Augmented LLMs

Reinforcement learning (RL) improves agent performance by learning from previous experience and maximizing cumulative rewards (Sutton & Barto, 1998). Recently, RL with group-based methods such as GRPO (Shao et al., 2024) have enabled LLMs to solve complex tasks such as Olympiad-level math problems as well as broader reasoning tasks (Shao et al., 2024; Yu et al., 2025; Wen et al., 2026). On the search side, retrieval-augmentation improves LLM by integrating external knowledge (Khandelwal et al., 2020; Lewis et al., 2020; Gao et al., 2023). Furthermore, iteratively retrieving relevant documents improves LLM performance on complex questions (Jiang et al., 2023; Trivedi et al., 2023; Mallen et al., 2023; Yoran et al., 2024). In search agents, search engines are popularly incorporated as optional tools, leading to enhanced multi-hop reasoning performance, such as Search-R1, Search-o1, among many others (Jin et al., 2025; Song et al., 2025; Li et al., 2025b; Zheng et al., 2025b). We provide a comprehensive summary of search agents in Appendix A.

2.2 Comparison with Existing Search Agent Training Datasets

Designing complex user queries with verifiable answers for deep research is nontrivial, because these tasks inherently require extensive search and exploration to find the correct answer. This limits the scale of human-annotated evaluation data, since verifying correctness and grounding answers to evidence documents is time-consuming. Synthetic data generation with large language models offers a way to scale training data, as shown in Table 1. Available datasets such as NQ (Kwiatkowski et al., 2019) and HotpotQA (Yang et al., 2018), commonly used for finetuning search agents, such as Search-R1, contain easy queries, solvable within 1–2 hops of search.

This motivated works on constructing deep search datasets with complex queries. Existing works such as DeepDive (Lu et al., 2025), WebShaper (Tao et al., 2026) and WebSailor (Li et al., 2025a) typically rely on hyperlink structure or knowledge graphs to synthesize queries from graph context. Similarly, works such as MoNaCo (Wolfson et al., 2026) and InForage (Qian & Liu, 2025) rely on human annotators, whereas InfoSeek (Xia et al., 2026), a concurrent effort constructs synthetic datasets via entity linking and obfuscation. Nevertheless, existing methods primarily include cumbersome pre-requisites that inherently limit the reproduction or real-world adoption. In contrast, we target a strict *no prerequisites* setting, relying exclusively on search-enabled LLMs to provide us synthetic QA training pairs with reasonable complexity (an example in Table 2) without requiring any paid API services. In terms of scale, ORBIT contains 20K training pairs, the largest amongst concurrent training datasets containing queries with hard complexity, covering both Wikipedia and Web (Britannica, NIH, StackExchange, etc.) as sources on 15 diverse domains.

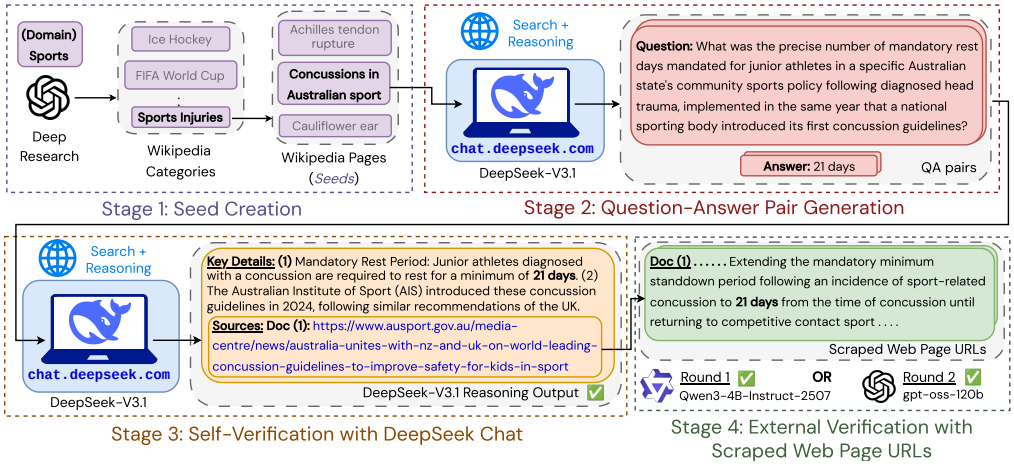

Figure 1: An end-to-end data instance in the **ORBIT** dataset generated using our framework. The procedure involves the following stages: (1) *seed creation*, (2) *question–answer pair generation*, (3) *self-verification with DeepSeek Chat*, and (4) *external verification with scraped web page URLs*. Each stage is described in detail in Section 3.1.

3 ORBIT: A Scalable and Verifiable Training Dataset for Search Agents

ORBIT consists of 20K *reasoning-intensive* or complex queries, often requiring *multi-hops* of searching information to be able to answer them confidently. We carefully constructed the dataset through a fully-automatic four-stage framework, using seeds constructed across 15 diverse domains. We first describe the data collection pipeline in Section 3.1, then analyze the dataset statistics in Section 3.2.

3.1 ORBIT Dataset Creation

ORBIT dataset creation requires zero pre-requisites allowing researchers to easily adopt the framework. Each training pair in ORBIT includes an *inverted* question and *short verifiable answer* (example in Table 2), where it is easier to verify the question given the answer (Wei et al., 2025). The dataset construction framework comprises four stages, described below:

Stage 1: Seed Creation. We manually decide on a set of 15 broad domains to be used, 12 of these broad domains are inspired from BrowseComp (Wei et al., 2025): *TV Shows & Movies*, *Science & Technology*, *Art*, *History*, *Sports*, *Music*, *Video Games*, *Geography*, *Politics*, *Medicine*, *Finance* and *Law*. We add three more: *Puzzles*, *Mathematics* and *Code*. Next, using OpenAI's `Deep Research` (`chat.openai.com`), we generate at least 100 different Wikipedia categories for each domain that are important, with a focus on diversity covering both frequent (head) and tail-end categories. We defer to the complete list of categories for each domain in Table 13. Furthermore, we use MediaWiki's REST API[1] to retrieve Wikipedia pages that are linked to a particular Wikipedia category. We use the linked Wikipedia page titles as **seeds** during question–answer pair generation in the next stage.

Stage 2: Question-Answer Pair Generation. We utilize the Wikipedia page title as the seed as an inspiration to generate an *inverted* question, that is *reasoning-intensive*, requiring multi-hop verifications to reach the generated answer. Specifically, we utilize `DeepSeek-V3.1` by sending each prompt on DeepSeek's Chat GUI (`chat.deepseek.com`) automatically using Python Selenium. We require a manual login to the website once, and activate the `DeepThink` & `Search` buttons. `DeepSeek-V3.1` conducts a web-search, gathers information, reasons and thinks carefully by checking each information hop, and generates the question–answer pair in the required output format. The prompt includes a single-shot examplar (shuffled for

[1] `https://www.mediawiki.org/wiki/API:REST_API`

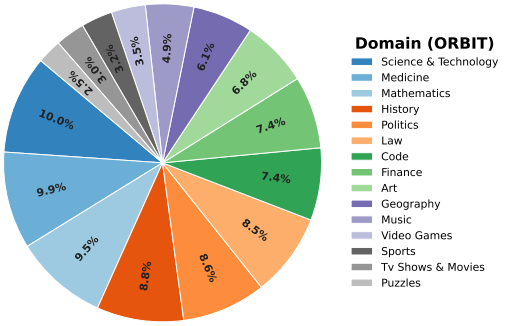

Figure 2a: **ORBIT domain composition.** We have question–answer pairs distributed evenly across 15 domains. The highest number of pairs come from *Science & Technology* (10%) and lowest from *Puzzles* (2.5%).

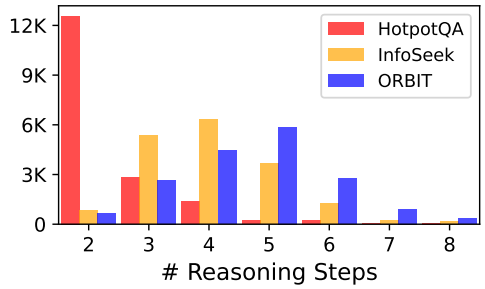

Figure 2b: **Question complexity.** ORBIT contains complex or deep search queries requiring on average 4–6 reasoning steps, that is higher than existing datasets: InfoSeek (3–5 steps) and HotpotQA (2 steps).

answer diversity[2]) and is provided in Figure 6. We generated a total of 44.1K training pairs from Stage 2 with `DeepSeek-V3.1`, requiring 2–3 months for the completion of this stage.

Stage 3: Self-Verification with DeepSeek Chat. After generating synthetic training pairs from the previous stage, we conduct *self-verification* for the question–answer pair. For this stage, we require a high-performing LLM with access to web-search and closed-sourced search APIs are expensive. Therefore, to limit costs, we independently prompt the same model: `DeepSeek-V3.1` from the DeepSeek Chat GUI (`chat.deepseek.com`) with enabled `DeepThink` and `Search` buttons, by providing the generated question and the answer as input, and ask `DeepSeek-V3.1` to reason and verify whether the answer satisfies all criteria and sub-parts mentioned in the question. The self-verification prompt is shown in Figure 7. DeepSeek Chat first conducts a web-search by searching a maximum of 50 documents (before starting to think) to be able to verify and cite the required sources as verification statements. Once completed, we provide the reasoning output to `Qwen3-4B-Instruct-2507`[3] to filter out training pairs with answers without sufficient verification, i.e., fully incorrect, or those only partially answered the question. After Stage 3, we were able to self-verify 61.5% of all training pairs from Stage 2, remaining with 27.1K training pairs.

Stage 4: External Verification with Scraped Web Page URLs. To avoid self bias using DeepSeek-V3.1 for both pair generation and verification (Panickssery et al., 2024), we utilize two seperate LLMs in a cascading setup to conduct an external verification. We scrape the web context available from the web page URLs provided during the self-verification stage using Python Selenium, and parse the document with Trafilatura (Barbaresi, 2021). In the first round, we prompt the `Qwen3-4B-Instruct-2507` judge by providing the question, answer and scraped web page document information, and ask the LLM to independently reason and provide the answer. Next, we use the same LLM to evaluate whether the LLM judge answer and the gold answer match. The external verification prompt is provided in Figure 8. For all unmatched pairs in the first round, we repeat the procedure with the OpenAI's `gpt-oss-120b`[4] as the judge in the next round. After external verification from both rounds, we verify 19,790 training pairs. Lastly, the authors manually verify and add 210 training pairs by checking the non-verified 400–500 training pairs by interacting with OpenAI's ChatGPT (`chat.openai.com`) (GPT-5.2 with `Search` button enabled).

3.2 ORBIT Dataset Statistics

The ORBIT dataset statistics are provided in Table 3. Generated questions are long, containing **64 tokens** on average and answers are short (and verifiable) with **3.5 tokens** (measured

[2]We empirically observed that shuffling the exemplar helped generate question–answer pairs with diverse types of answers, e.g., not always fixated on the year or a numerical answer.

[3]`https://huggingface.co/Qwen/Qwen3-4B-Instruct-2507`

[4]`https://huggingface.co/openai/gpt-oss-120b`

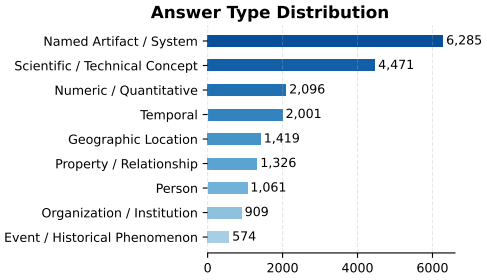

Figure 3a: **Answer type distribution.** Answers are a mixture of named artifacts, scientific or technical concepts, or even historical phenomenon. Answer type definitions are provided in Table 9.

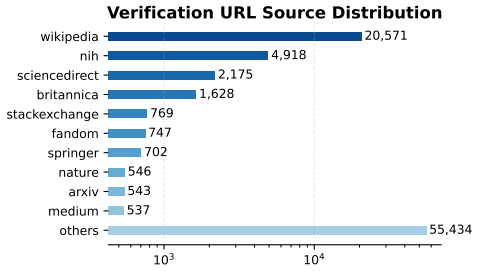

Figure 3b: **URL source distribution.** Verification URL's from the head distribution include Wikipedia, NIH, Britannica among others. There is a long-tail of source distribution.

using the OpenAI's tiktoken package). The average number of estimated reasoning steps (by counting the number of verification statements) is 4.42, thereby requiring 4–5 information hops to verify the answer. Similarly, there are about 4.36 verification URLs on average, with almost 3.35× of non-Wikipedia URLs over Wikipedia URLs.

Domain Composition. We show the domain composition present in ORBIT in Figure 2a. Questions from each domain in ORBIT are generated uniformly and filtered by removing unverified question–answer pairs. *Science & Technology* has the highest composition of question–answer pairs with 10% and *Puzzles* has the lowest composition with 2.5%. To summarize, ORBIT contains question–answer pairs spread evenly across all domains.

Question Complexity. We evaluate the amount of information-seeking or reasoning hops necessary to faithfully answer each part of the question. We work on three datasets: (1) HotpotQA (Yang et al., 2018), (2) InfoSeek (Xia et al., 2026) and ORBIT. To avoid the uneven distribution of training pairs, we randomly sample 18K training pairs from each dataset. We decompose the original question into necessary sub-questions with Qwen3-8B (Yang et al., 2025a) (with thinking mode enabled), and count each sub-question as a separate information hop *necessary* to verify a part of the question. Our results in Figure 2b, show that ORBIT contains the most complex queries, requiring on average 4–6 reasoning steps to answer them. In contrast, InfoSeek contains 3–5 reasoning steps and HotpotQA with 2 reasoning steps.

Dataset Statistic	Value
# Question–Answer pairs	20,147
- Avg. # question tokens	63.87
- Avg. # answer tokens	3.46
- Avg. # reasoning steps	4.42
Avg. Verification URLs	4.36
- Non-Wikipedia URLs	3.35
- Wikipedia URLs	1.02

Table 3: **ORBIT dataset statistics**. URL statistics reflect the distinct sources required for answer verification.

Answer Type and URL Source Distribution. In Figure 3a, we automatically classify the short factual answers in ORBIT into each answer type category using Qwen3-8B (with thinking mode enabled). In ORBIT, we observe that a majority of answers are either a named or system artifact, or a scientific or technical concept. We observe 20% of answers based on temporal entity (e.g., calendar date, year etc.) or numeric (numerical value, measurement etc). Overall, the plot shows the diversity in ORBIT answers. Similarly, we also plot the URL source distribution to identify major URL sources required for answer verification. From Figure 3b, we observe that Wikipedia is the largest source, with NIH and ScienceDirect, being the second and third-largest sources. There is a long-tail distribution of source domains (Others is the largest category) necessary for answer verification.

4 Experimental Setup

GRPO Training Details. We use the `verl-tool` repository (Jiang et al., 2025) for training search agents using GRPO (Shao et al., 2024). We use Qwen3-4B[5] (Yang et al., 2025a) as

[5] https://huggingface.co/Qwen/Qwen3-4B

Model	Single-Hop QA			Multi-Hop QA				Avg. 7
	NQ	TQA	PopQA	HQA	2Wiki	MSQ	Bamb	
Search Agents (Base: Qwen2.5-3B-Instruct)								
Search-o1-3B	23.8	48.2	26.2	22.1	21.8	5.4	32.0	25.6
Search-R1-3B	40.8	59.1	42.8	30.8	31.1	8.4	13.0	32.3
ZeroSearch-3B	41.2	61.5	44.0	31.2	33.2	12.6	14.3	34.0
AutoRefine-3B	43.6	59.7	44.7	40.4	38.0	16.9	33.6	39.6
InForage-3B	42.1	59.7	45.2	40.9	42.8	17.2	36.0	40.6
InfoSeeker-3B	42.7	57.1	48.0	**44.6**	52.0	**20.5**	39.8	43.5
Search Agents (Base: Qwen3-4B)								
Search-R1-4B†	38.5	60.1	46.8	34.2	46.4	13.2	41.6	40.1
InfoSeeker-4B†	37.0	64.1	47.0	39.8	59.0	15.6	51.2	44.8
ORBIT-4B†	**43.7**	**67.3**	**53.8**	42.5	**61.1**	20.1	**55.2**	**49.1**

Table 4: Performance comparison on single-hop and multi-hop QA benchmarks of search agents <4B parameters. Best scores are highlighted in **bold**. (†) denotes models that were trained in our work using Qwen3-4B as the base model. In addition, EM accuracy results using the 2018 Wikipedia corpus and E5-base-v2 retriever are provided in Table 8.

the base model with the following hyperparameters: rollouts = 8, maximum turns = 5, maximum response length = 8192, learning rate = 1e-6, train batch size = 512, KL loss coefficient = 0.001, temperature = 1.0 and top-p = 1.0. The reward function is an exact match of the ground truth answer, identical to Search-R1 (Jin et al., 2025). ORBIT-4B have been trained for 160 training steps, utilizing 4xH100 GPUs for 1–2 days. Importantly, we mix ORBIT, NQ and HotpotQA in an equal mixture (1: 1: 1), as we observe that non-complex training pairs with 1–2 reasoning steps (i.e., NQ and HotpotQA) are crucial as a learning step. Additional details including the prompt template and full hyperparameter choices are provided in Appendix B and Appendix C.

Training Search Tool. For the search tool, we adopt the **D**ux **D**istributed **G**lobal **S**earch or **DDGS** framework, an open and free metasearch library that aggregates results from diverse web search services, including Google, Brave, Bing etc.[6] The framework returns the title and a relevant snippet of the web document. Due to budget restrictions, we are unable to use paid web search tools in our work, such as Google Search API, Exa or SerpAPI.

Evaluation Setup & Datasets. We embed the June 2024 Wikipedia dump where articles are split into passages[7] and retrieve top-5 passages using the BGE-M3 retrieval model (Chen et al., 2024), replicating the evaluation setting in **InfoSeeker** (Xia et al., 2026). We evaluate on three single-hop Wikipedia QA benchmarks: Natural Questions (NQ) (Kwiatkowski et al., 2019), TriviaQA (TQA) (Joshi et al., 2017) and PopQA (Mallen et al., 2023), and four multi-hop Wikipedia QA benchmarks: HotpotQA (HQA) (Yang et al., 2018), 2WikiMultihopQA (2Wiki) (Ho et al., 2020b), MuSiQue (MSQ) (Trivedi et al., 2022), and Bamboogle (Bamb) (Press et al., 2023). In addition, we include FRAMES (Krishna et al., 2025) for a challenging multi-hop evaluation. We use Exact Match (EM) accuracy as the evaluation metric for all datasets.

Search Agent Baselines. A majority of existing search agent baselines rely on Qwen2.5 variants as the base model (Yang et al., 2025b). As prior work does not provide sufficient details on the training setup, data or code, we are unable to reproduce them with Qwen3-4B. Therefore, to enable a fair comparison, we gather numbers from InfoSeek (Xia et al., 2026), with an identical setup as ours, in terms of the corpus, search tool model and configuration. Next, we demonstrate the effectiveness of the ORBIT dataset, by training two comparable search agent baselines using Qwen3-4B as the base model (Yang et al., 2025a):

[6] https://github.com/deedy5/ddgs

[7] https://huggingface.co/datasets/Upstash/wikipedia-2024-06-bge-m3

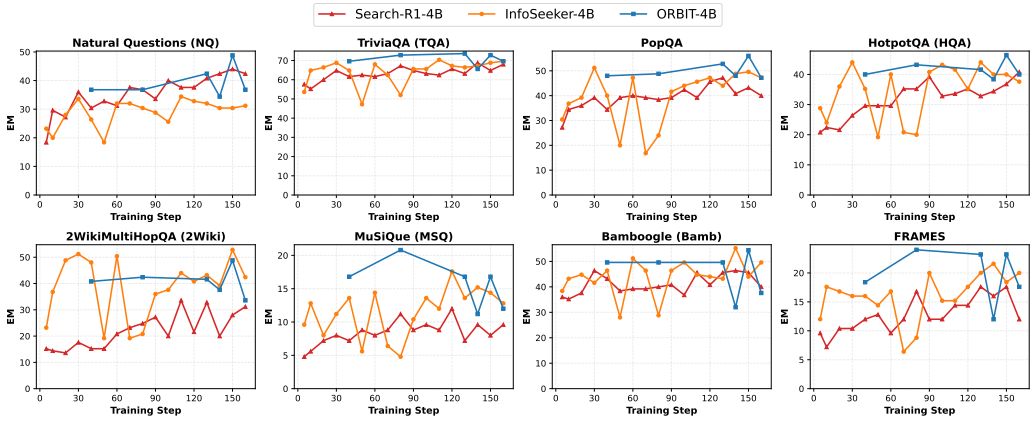

Figure 4: Validation EM accuracy of Search-R1-4B, InfoSeeker-4B and ORBIT-4B on 160 training steps on Wikipedia datasets, each with 125 randomly sampled validation pairs. The accuracy drops observed during training primarly occurs due to DDGS or live web retriever, that can potentially downgrade search results when servers are busy (e.g., `google` → `bing`).

1. **Search-R1-4B**: a reproduction of Search-R1 (Jin et al., 2025), trained using GRPO with the Qwen3-4B base model on the complete NQ and HotpotQA dataset.
2. **InfoSeeker-4B**: a reproduction of InfoSeeker (Xia et al., 2026), trained using GRPO with the Qwen3-4B base model on an equal mixture (1: 1: 1) of NQ, HotpotQA and InfoSeek.

5 Experimental Results & Ablations

Overall Performance. We first discuss the main Wikipedia evaluation results in Table 4. Using Qwen3-4B as the base model, ORBIT-4B consistently outperforms Search-R1-4B and InfoSeeker-4B on all single-hop and multi-hop QA Wikipedia datasets, by achieving 49.1 EM accuracy on average, that is 9.0 points better than Search-R1-4B and 4.3 points better than InfoSeeker-4B. This shows the importance of the ORBIT dataset over traditional datasets such as NQ and HotpotQA, and similar reasoning-intensive datasets such as InfoSeek. We observe similar gains with ORBIT-4B in the Search-R1 evaluation setup in Appendix D and in the validation EM accuracy scores on all datasets provided in Figure 4.

Comparison with Qwen2.5-3B-Instruct as the Base Model. From Table 4, ORBIT-4B is compared with several baselines trained using the Qwen2.5-3B-Instruct as the base model, taken from InfoSeek (Xia et al., 2026). We observe that direct GRPO fine-tuning on a newer and recent base model (e.g., Qwen3-4B), significantly outperforms search agents using Qwen2.5-3B-Instruct as the base model. Supervised fine-tuning (SFT) is beneficial for the Qwen2.5 as the base model as an alignment stage. However, GRPO training is sufficient for training Qwen3-4B as the base model. InfoSeeker-3B is the strongest Qwen2.5-3B-Instruct baseline achieving 43.5 EM score on average, 5.6 points below ORBIT-4B.

Extended Results on FRAMES. We extend our evaluation on FRAMES (Krishna et al., 2025), a recent and challenging multi-hop Wikipedia QA dataset. From Table 5, we observe that ORBIT-4B achieves the highest EM accuracy in contrast to Search-R1-4B and InfoSeeker-4B. Although, HotpotQA, NQ and InfoSeek are all in-domain, i.e., Wikipedia datasets, fine-tuning on ORBIT (Wikipedia and Web) helps in answering multi-hop reasoning questions in FRAMES.

Model	FRAMES
Search-R1-4B†	14.7
InfoSeeker-4B†	18.5
ORBIT-4B†	**24.2**

Table 5: EM accuracy on FRAMES.

Dataset Mixing Ratios. We inspect how dataset composition affects the EM accuracy of ORBIT-4B. Specifically, we vary the distribution of the training pairs from three datasets,

ORBIT-4B	Single-Hop QA			Multi-Hop QA					Avg. 8
	NQ	TQA	PopQA	HQA	2Wiki	MSQ	Bamb	FRAMES	
Dataset Mixing Ratios (Dataset ratio of NQ: HotpotQA: ORBIT)									
Ratio (1: 1: 1)	**43.7**	**67.3**	**53.8**	**42.5**	61.1	**20.1**	**55.2**	**24.2**	**46.0**
Ratio (1: 2: 4)	41.7	65.3	50.8	40.0	**61.3**	16.1	52.0	19.1	43.3
Top #K Search Results (Snippet & title only during training)									
Top 5	43.7	**67.3**	53.8	42.5	**61.1**	20.1	**55.2**	**24.2**	**46.0**
Top 10	**43.9**	66.0	**55.2**	**43.1**	58.4	**21.0**	51.2	21.4	45.0

Table 6: Ablations on dataset mixing ratios and top-k search results with ORBIT-4B.

NQ: HotpotQA: ORBIT at (1: 2: 4) and our default (1: 1: 1). The (1: 2: 4) dataset mixing ratio reduces the distribution of the single-hop pairs and provides more focus on multi-hop training pairs. These comparative results are detailed in Table 6. The optimal setting is the equal dataset mix (1: 1: 1), achieving 46.0 EM accuracy, on average. This suggests that for smaller search agents, training directly on challenging multi-hop questions isn't the optimal strategy. Rather training on an equally mixed training dataset with easy single-hop or multi-hop questions present in Natural Questions and HotpotQA, helps to teach search agents effectively on answering questions requiring multi-hop steps of information.

Top-K Search Results. Further, we also inspect how the top-k DDGS search results during each turn affect the EM accuracy of ORBIT-4B. As shown in Table 6, the results are counterintuitive, as more search results doesn't correlate to better EM accuracy. In ORBIT-4B, retrieving top 5 search results offers a better trade-off in terms of search latency. We suspect that the search agent observes relevant titles and snippets by avoiding additional search distractors, which lowers the EM accuracy marginally by 1.0 point, on average.

6 Discussion

Dataset Generation Under Budget Constraints. An important objective is the ability to fully open-source and generate a synthetic training dataset under a tight budget. However, the applicability of the framework is not restricted to a frugal setup. With larger budgets, the framework can be constructed more reliably with paid access to capable LLMs and search APIs, reducing latency to construct faster and better quality synthetic training datasets. In this work, we demonstrate that even with minimal cost and a single laptop, a competitive and reasoning-intensive synthetic training dataset for search agents can be generated.

Training Setup Limitations. Due to budget constraints, we focus on small search agents and a relatively minimal training setup. Our GRPO training setup is sufficient to demonstrate the effectiveness of ORBIT over existing baselines with comparable resources. However, to close the gap with larger open-source and proprietary search agents, several extensions would be necessary: (1) access to paid search APIs; (2) an additional `get_document` tool that fetches and summarizes the full content of a retrieved webpage; and (3) a larger base model ($\geq$ 30B parameters) combined with SFT warm-up followed by RL fine-tuning (Xia et al., 2026; Qian & Liu, 2025). We leave these directions as future work.

7 Conclusion

We introduce ORBIT, a scalable and verifiable dataset with 20K reasoning-intensive question–answer pairs for training search agents on a tight budget. The dataset is built with a four-stage framework that combines seed construction, question–answer pair generation, self-verification, and external verification, without relying on paid search APIs or expensive human annotation. Using ORBIT, we train ORBIT-4B, a search agent using Qwen3-4B as the base model optimized with GRPO. ORBIT-4B achieves strong performance among

sub-4B search agents and consistently outperforms the baselines on single-hop and multi-hop Wikipedia QA datasets. Overall, our findings demonstrate that careful synthetic data generation is a viable path for improving search agents without proprietary tools or large annotation budgets. We hope ORBIT lowers the barrier to research on deep search and motivates future work on open, reproducible data pipelines for open-source search agents.

Acknowledgments

We begin by acknowledging the old but resilient Nandan's 2018 Linux laptop used to construct the ORBIT dataset. We sincerely thank DeepSeek for providing an accesible chat interface with integrated web search capabilities. We are also grateful to the Digital Research Alliance of Canada for providing access to the Nibi & Fir clusters (only clusters with internet access) used for training search agents with 4xH100's. This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Ethics Statement

ORBIT is constructed from publicly accessible sources, primarily Wikipedia categories, publicly retrievable and sourced webpages, and generated question–answer pairs. We do not intentionally collect private, sensitive, or personally identifying information, and the dataset is intended for factual question answering research rather than profiling individuals.

Stages 2 and 3 of the ORBIT framework rely on automated interactions with Python Selenium using the DeepSeek Chat web interface (`chat.deepseek.com`). To ensure an ethical practice, we took the following steps: First, all interactions are performed on DeepSeek Chat are using the publicly accesible web interface under a *single authenticated personal account*, consistent with how a user would use the chat service manually. Second, the automation is *rate-limited* and kept sequential, where each prompt is submitted *one at a time* and the Selenium script waits for a complete model response before proceeding for the next one. Third, the purpose of the dataset construction is *purely academic research* with no commercial intent, and the generated content of the dataset will be licensed under the CC BY-NC-SA, to avoid commercial usage. Fourth, the study follows the DeepSeek Terms of Use[8], where it states that inputs and outputs from the DeepSeek Chat service *can be applied towards academic research*. We will explicitly highlight these artifacts with our dataset release; researchers who adopt our dataset should review the applicable terms of any service they use and seek guidance where appropriate.

Reproducibility Statement

We aim to make the dataset construction process as reproducible as possible. Due to budget constraints, we use the free and open-source DDGS search aggregator that reduces search reproducibility in our work, as the search results can vary based on the DDGS server load. Similarly, at the time of ORBIT dataset construction, we interact with `DeepSeek-V3.1` utilized in DeepSeek Chat, however, the models continue to upgrade and the exact model might not be available at a future date, leading to change in outputs. To mitigate this, we are fully open-sourcing and releasing intermediate artifacts wherever possible, including question–answer pairs, cited URLs, verification metadata and experiment configurations, to enable other researchers that can easily construct synthetic training data with limited access to budget, making it a lot more accesible to a wider community.

References

Adrien Barbaresi. Trafilatura: A web scraping library and command-line tool for text discovery and extraction. In Heng Ji, Jong C. Park, and Rui Xia (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International*

[8] https://cdn.deepseek.com/policies/en-US/deepseek-terms-of-use.html

Joint Conference on Natural Language Processing: System Demonstrations, pp. 122–131, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-demo.15. URL https://aclanthology.org/2021.acl-demo.15/.

Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. M3-embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2318–2335, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.137. URL https://aclanthology.org/2024.findings-acl.137/.

Mingyang Chen, Linzhuang Sun, Tianpeng Li, sunhaoze, ZhouYijie, Chenzheng Zhu, Haofen Wang, Jeff Z. Pan, Wen Zhang, Huajun Chen, Fan Yang, Zenan Zhou, and Weipeng Chen. ReSearch: Learning to reason with search for LLMs via reinforcement learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=OuGAwwAT8G.

Tianyu Gao, Howard Yen, Jiatong Yu, and Danqi Chen. Enabling large language models to generate text with citations. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 6465–6488, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.398. URL https://aclanthology.org/2023.emnlp-main.398/.

Gemini. Gemini Deep Research. 2025. URL https://gemini.google/overview/deep-research/.

Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara, and Akiko Aizawa. Constructing a multi-hop QA dataset for comprehensive evaluation of reasoning steps. In Donia Scott, Nuria Bel, and Chengqing Zong (eds.), *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6609–6625, Barcelona, Spain (Online), December 2020a. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.580. URL https://aclanthology.org/2020.coling-main.580/.

Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara, and Akiko Aizawa. Constructing a multi-hop QA dataset for comprehensive evaluation of reasoning steps. In Donia Scott, Nuria Bel, and Chengqing Zong (eds.), *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6609–6625, Barcelona, Spain (Online), December 2020b. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.580. URL https://aclanthology.org/2020.coling-main.580/.

Dongfu Jiang, Yi Lu, Zhuofeng Li, Zhiheng Lyu, Ping Nie, Haozhe Wang, Alex Su, Hui Chen, Kai Zou, Chao Du, Tianyu Pang, and Wenhu Chen. VerlTool: Towards Holistic Agentic Reinforcement Learning with Tool Use, 2025. URL https://arxiv.org/abs/2509.01055.

Zhengbao Jiang, Frank Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yiming Yang, Jamie Callan, and Graham Neubig. Active retrieval augmented generation. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 7969–7992, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.495. URL https://aclanthology.org/2023.emnlp-main.495/.

Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan O Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-R1: Training LLMs to reason and leverage search engines with reinforcement learning. In *Second Conference on Language Modeling*, 2025. URL https://openreview.net/forum?id=Rwhi91ideu.

Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. In Regina Barzilay and Min-Yen Kan (eds.), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1601–1611, Vancouver, Canada,

July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1147. URL https://aclanthology.org/P17-1147/.

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

Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. Generalization through memorization: Nearest neighbor language models. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HklBjCEKvH.

Satyapriya Krishna, Kalpesh Krishna, Anhad Mohananey, Steven Schwarcz, Adam Stambler, Shyam Upadhyay, and Manaal Faruqui. Fact, fetch, and reason: A unified evaluation of retrieval-augmented generation. In Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 4745–4759, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.naacl-long.243. URL https://aclanthology.org/2025.naacl-long.243/.

Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. Natural questions: A benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:452–466, 2019. doi: 10.1162/tacl_a_00276. URL https://aclanthology.org/Q19-1026/.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.

Jian Li, Dongsheng Chen, Zhenhua Xu, Yizhang Jin, Jiafu Wu, Chengjie Wang, Xiaotong Yuan, and Yabiao Wang. Improving search agent with one line of code, 2026a. URL https://arxiv.org/abs/2603.10069.

Jian Li, Yizhang Jin, Dongqi Liu, Hang Ding, Jiafu Wu, Dongsheng Chen, Yunhang Shen, Yulei Qin, Ying Tai, Chengjie Wang, Xiaotong Yuan, and Yabiao Wang. SE-Search: Self-evolving search agent via memory and dense reward, 2026b. URL https://arxiv.org/abs/2603.03293.

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, 2025a. 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 Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 5420–5438, Suzhou, China, November 2025b. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.276. URL https://aclanthology.org/2025.emnlp-main.276/.

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.

Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 9802–9822, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.546. URL https://aclanthology.org/2023.acl-long.546/.

Jianbiao Mei, Tao Hu, Daocheng Fu, Licheng Wen, Xuemeng Yang, Rong Wu, Pinlong Cai, Xinyu Cai, Xing Gao, Yu Yang, Chengjun Xie, Botian Shi, Yong Liu, and Yu Qiao. O^2-searcher: A searching-based agent model for open-domain open-ended question answering, 2025. URL https://arxiv.org/abs/2505.16582.

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

Arjun Panickssery, Samuel R. Bowman, and Shi Feng. LLM evaluators recognize and favor their own generations. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=4NJBV6Wp0h.

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 Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), *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/.

Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt, Noah A. Smith, and Mike Lewis. Measuring and narrowing the compositionality gap in language models, 2023. URL https://openreview.net/forum?id=PUwbwZJz9dO.

Hongjin Qian and Zheng Liu. Scent of knowledge: Optimizing search-enhanced reasoning with information foraging. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=26kUrQm4zw.

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, 2024. URL https://arxiv.org/abs/2402.03300.

Yaorui Shi, Sihang Li, Chang Wu, Zhiyuan Liu, Junfeng Fang, Hengxing Cai, An Zhang, and Xiang Wang. Search and refine during think: Facilitating knowledge refinement for improved retrieval-augmented reasoning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=rBlWKIUQey.

Zhengliang Shi, Lingyong Yan, Dawei Yin, Suzan Verberne, Maarten de Rijke, and Zhaochun Ren. Iterative self-incentivization empowers large language models as agentic searchers. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=s9NkfkUuEr.

Huatong Song, Jinhao Jiang, Yingqian Min, Jie Chen, Zhipeng Chen, Wayne Xin Zhao, Lei Fang, and Ji-Rong Wen. R1-Searcher: Incentivizing the Search Capability in LLMs via Reinforcement Learning, 2025. URL https://arxiv.org/abs/2503.05592.

Hao Sun, Zile Qiao, Jiayan Guo, Xuanbo Fan, Yingyan Hou, Yong Jiang, Pengjun Xie, Yan Zhang, Fei Huang, and Jingren Zhou. ZeroSearch: Incentivize the Search Capability of LLMs without Searching, 2025a. URL https://arxiv.org/abs/2505.04588.

Shuang Sun, Huatong Song, Yuhao Wang, Ruiyang Ren, Jinhao Jiang, Junjie Zhang, Fei Bai, Jia Deng, Wayne Xin Zhao, Zheng Liu, Lei Fang, Zhongyuan Wang, and Ji-Rong Wen. SimpleDeepSearcher: Deep information seeking via web-powered reasoning trajectory synthesis. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 13705–13720, Suzhou, China, November 2025b. Association for Computational Linguistics. ISBN 979-8-89176-335-7. doi: 10.18653/v1/2025.findings-emnlp.739. URL https://aclanthology.org/2025.findings-emnlp.739/.

R.S. Sutton and A.G. Barto. Reinforcement learning: An introduction. *IEEE Transactions on Neural Networks*, 9(5):1054–1054, Sep. 1998. ISSN 1941-0093. doi: 10.1109/TNN.1998.712192.

Zhengwei Tao, Jialong Wu, Wenbiao Yin, Pu Wu, Junkai Zhang, Baixuan Li, Haiyang SHEN, Kuan Li, Liwen Zhang, Xinyu Wang, Wentao Zhang, Yong Jiang, Pengjun Xie, Fei Huang, and Jingren Zhou. WebShaper: Agentically Data Synthesizing via Information-Seeking Formalization. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=hld4TzJsnD.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. MuSiQue: Multihop questions via single-hop question composition. *Transactions of the Association for Computational Linguistics*, 10:539–554, 05 2022. ISSN 2307-387X. doi: 10.1162/tacl_a_00475. URL https://doi.org/10.1162/tacl_a_00475.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10014–10037, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.557. URL https://aclanthology.org/2023.acl-long.557/.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training, 2024. URL https://arxiv.org/abs/2212.03533.

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.

Xumeng Wen, Zihan Liu, Shun Zheng, Shengyu Ye, Zhirong Wu, Yang Wang, Zhijian Xu, Xiao Liang, Junjie Li, Ziming Miao, Jiang Bian, and Mao Yang. Reinforcement learning with verifiable rewards implicitly incentivizes correct reasoning in base LLMs. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=jGbRWwIidy.

Tomer Wolfson, Harsh Trivedi, Mor Geva, Yoav Goldberg, Dan Roth, Tushar Khot, Ashish Sabharwal, and Reut Tsarfaty. MoNaCo: More natural and complex questions for reasoning across dozens of documents. *Transactions of the Association for Computational Linguistics*, 14:23–46, 2026. doi: 10.1162/tacl.a.64. URL https://aclanthology.org/2026.tacl-1.2/.

Jialong Wu, Wenbiao Yin, Yong Jiang, Zhenglin Wang, Zekun Xi, Runnan Fang, Linhai Zhang, Yulan He, Deyu Zhou, Pengjun Xie, and Fei Huang. WebWalker: Benchmarking LLMs in web traversal. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and

Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10290–10305, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.508. URL https://aclanthology.org/2025.acl-long.508/.

Yunjia Xi, Jianghao Lin, Menghui Zhu, Yongzhao Xiao, Zhuoying Ou, Jiaqi Liu, Tong Wan, Bo Chen, Weiwen Liu, Yasheng Wang, Ruiming Tang, Weinan Zhang, and Yong Yu. Infodeepseek: Benchmarking agentic information seeking for retrieval-augmented generation, 2025. URL https://arxiv.org/abs/2505.15872.

Ziyi Xia, Kun Luo, Hongjin Qian, Siqi Bao, and Zheng Liu. Open Data Synthesis For Deep Research. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=2c9TjRbAib.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025a. URL https://arxiv.org/abs/2505.09388.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 Technical Report, 2025b. URL https://arxiv.org/abs/2412.15115.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2369–2380, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1259. URL https://aclanthology.org/D18-1259/.

Ori Yoran, Tomer Wolfson, Ori Ram, and Jonathan Berant. Making retrieval-augmented language models robust to irrelevant context. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ZS4m74kZpH.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, YuYue, Weinan Dai, Tiantian Fan, Gaohong Liu, Juncai Liu, LingJun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Ru Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Yonghui Wu, and Mingxuan Wang. DAPO: An open-source LLM reinforcement learning system at scale. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=2a36EMSSTp.

Zhenrui Yue, Kartikeya Upasani, Xianjun Yang, Suyu Ge, Shaoliang Nie, Yuning Mao, Zhe Liu, and Dong Wang. Dr. Zero: Self-Evolving Search Agents without Training Data, 2026. URL https://arxiv.org/abs/2601.07055.

Yaocheng Zhang, Haohuan Huang, Zijun Song, Yuanheng Zhu, Qichao Zhang, Zijie Zhao, and Dongbin Zhao. CriticSearch: Fine-Grained Credit Assignment for Search Agents via a Retrospective Critic, 2025. URL https://arxiv.org/abs/2511.12159.

Xuhui Zheng, Kang An, Ziliang Wang, Yuhang Wang, and Yichao Wu. StepSearch: Igniting LLMs search ability via step-wise proximal policy optimization. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 21805–21830, Suzhou, China, November 2025a. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.1106. URL https://aclanthology.org/2025.emnlp-main.1106/.

Yuxiang Zheng, Dayuan Fu, Xiangkun Hu, Xiaojie Cai, Lyumanshan Ye, Pengrui Lu, and Pengfei Liu. DeepResearcher: Scaling deep research via reinforcement learning in real-world environments. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 414–431, Suzhou, China, November 2025b. Association for Computational Linguistics. ISBN 979-8-89176-332-6. doi: 10.18653/v1/2025.emnlp-main.22. URL https://aclanthology.org/2025.emnlp-main.22/.

A Extended Related Work

This section contains a detailed overview of related search agent papers:

Search-o1 (Li et al., 2025b) One of the first works to incorporate multi-turn retrieval and reasoning by converting the retrieved documents into focused reasoning steps that integrate the external knowledge, while maintaining the logical flow of the reasoning chain. The authors used the QwQ-32B-Preview as the base model, employed the Bing Web Search for web search and Jina Reader API to fetch the content of the web page. The QwQ-32B-Preview was not fine-tuned using reinforcement learning, instead was naively used for multi-turn retrieval and reasoning on open-domain question-answering tasks on Wikipedia.

Search-R1 (Jin et al., 2025) One of the earliest works of training LLMs with multi-turn search engine calls and LLM response rollout using reinforcement learning with PPO. Search-R1 used two LLMs as the base model: Qwen2.5-3B/7B (Base/Instruct) (Yang et al., 2025b) and trained them on two Wikipedia datasets: NQ (Kwiatkowski et al., 2019) and HotpotQA (Yang et al., 2018). Both datasets contain queries that can be answered using 1–2 hops of information searches, the models are trained using a single RL stage containing a maximum of 2 turns and rollout 5 responses per prompt (GRPO). For evaluation, Search-R1 retrieved top 3 documents from the 2018 December Wikipedia corpus containing chunked documents with a maximum of 512 tokens (Karpukhin et al., 2020) with `e5-base-v2` (Wang et al., 2024)[9] as the dense retrieval model with 110M parameters, 12 layers and an embedding size of 768 dimensions. They observed PPO to be stable than GRPO for training models in their experiments.

ReSearch (Chen et al., 2025) ReSearch is similar to Search-R1. ReSearch trains Qwen2.5-7B/32B (Base/Instruct) LLMs using reinforcement learning with GRPO. ReSearch modified the Search-R1 reward by adding a format-based reward to teach the model to follow the required format. Models were trained on the training dataset of MuSiQue (Trivedi et al., 2022) with 5 rollouts. The evaluation setting is identical to Search-R1, with the only caveat of retrieving top-5 documents for each query.

R1-Searcher (Song et al., 2025) R1-Searcher is similar to both ReSearch and Search-R1. R1-Searcher trained Qwen2.5-7B (Base) and Llama-3.1-8B-Instruct LLMs using reinforcement learning with GRPO. They include a retrieval reward in the first-stage of RL training, followed by the answer reward in the second stage. They train model on two datasets: Hotpotqa (Yang et al., 2018) and 2WikiMultiHopQA (Ho et al., 2020a) with 16 rollouts. The evaluation setup is different from previous setups, they incorporate the English Wikipedia corpus provided by KILT (Petroni et al., 2021) and use the `BGE-large-en-v1.5`[10] as the dense retriever model, retrieving top-8 documents (except Bamboogle, where they employ Google web search API).

DeepResearcher (Zheng et al., 2025b) One of the first works of training LLMs with multi-turn search engine calls from a live search engine, such as Google web search API. They fine-tuned the Qwen2.5-7B Instruct model and trained on four datasets: NQ, TriviaQA, HotpotQA and 2WikiMultiHopQA with a distribution ratio of (1: 1: 3: 3) with a focus on multi-hop scenarios, using GRPO with 16 rollouts and a maximum turn size of 10. The evaluation was done with web search on a subset of 512 evaluation samples; making the reproducibility difficult, as the subsets are not publicly available.

SimpleDeepSearcher (Sun et al., 2025b) One of the first to show that SFT with reasoning trajectories works better than RL for training models. They SFT train four models: Qwen2.57B-Instruct, Qwen2.5-32B-Instruct, DeepseekDistilled-Qwen2.5-32B, and QwQ-32B with live web-search. Similar to DeepResearcher, they evaluate the models on 500 randomly sampled evaluation samples.

[9] https://huggingface.co/intfloat/e5-base-v2

[10] https://huggingface.co/BAAI/bge-large-en-v1.5

InForage (Qian & Liu, 2025) InForage introduces three reward mechanisms to incentivize comprehensive reasoning, by including an information gain reward, rewarding retrieval steps with valuable evidence and an efficiency penalty discouraging prolonged reasoning. InForage constructed their own dataset, where human annotators begin with a seed factoid claim and used Google Search API to extend the claim by looking at selected web-pages and expand the context around it. They manually annotated for 500 queries. Next, they expanded it automatically with GPT-4o for around 20K training samples. The Qwen2.5-Instruct 3B and 7B were used as the base models for training. Each model first went through a supervised fine-tuning stage (SFT) and further a reinforcement learning stage with PPO trained on a mixture of the self-constructed training dataset, NQ and HotpotQA. The maximum number of turns during training was 6, with cached Google search results and scraped full content of webpages. The inference setting is similar to Search-R1 except they use the `BGE-M3` model (Chen et al., 2024)[11] as the dense retrieval model.

InfoSeek (Xia et al., 2026) InfoSeek focuses on automated data synthesis, by constructing a training dataset to perform on reasoning-intensive queries using hierarchical constraints requiring raw entities and their relations. They fine-tune models using SFT via rejection sampling for the first stage of training followed by RL training in second stage. The train the Qwen2.5-3B-Instruct with 5 rollouts, and a maximum turn size of 10, with the search-engine retrieving top-5 contents. They fine-tune with GRPO on 17K harder samples (15K InfoSeek, 2K NQ and HotpotQA) in the second stage. The evaluation setting uses the Wikipedia corpus from 2025 segmented into chunks of 512 tokens and retrieve using the `BGE-M3` retrieval model with top-5 documents retrieved for each query.

WebSailor (Li et al., 2025a) WebSailor also focuses on automated data synthesis, by synthesizing high-uncertainty web navigation tasks to train specialized agents. WebSailor's approach is focused on building knowledge graphs and subgraph sampling and obfuscation. They train four models: Qwen2.5-3B/7B/32B/72B (Instruct) using SFT with rejection sampling and further train in the next stage using Duplicating Sampling Policy Optimization (DUPO) with 8 rollouts. No other training details were provided. They utilize two tools: (1) *search* for returning top-10 search results with Google Search, and (2) *visit* to summarize the information for each webpage using Qwen2.5-72B as the summary model.

WebShaper (Tao et al., 2026) WebShaper introduces a formalization-guided framework for data synthesis, by linking relationships between entities in order to expand existing questions. After constructing information seeking pairs, they use an expander agent to enhance the original question via iterative refinement. They train on three Qwen model variants: Qwen2.5-32B/72B and QwQ-32B using SFT with rejection sampling. Next, they train their models with DAPO with 8 rollouts. The temperature is set as 1.0, batch size as 128, and the learning rate as 1e-6.

B Search Agent Prompt

Figure 5 shows the prompt template used to train ORBIT-4B via GRPO in our work. The template instructs ORBIT-4B to first *plan* a solution trajectory by decomposing the question before issuing any search calls, a design choice motivated by the multi-hop, reasoning-intensive nature of ORBIT queries.

Concretely, the agent alternates between internal deliberation enclosed in `<think> </think>` tags and targeted web queries wrapped in `<search> </search>` tags. Each query is routed to DDGS web retrieval server (top-k=5 documents) whose results are injected back (containing the title and snippet information) into the context as `<information> </information>` blocks. The agent may issue arbitrarily many search calls per trajectory but is explicitly discouraged from repeating identical queries, which would otherwise dominate the turn budget without contributing new evidence. The trajectory concludes when the model emits a `<answer>`

[11] https://huggingface.co/BAAI/bge-m3

Search Agent Prompt

Answer the given question. Please break down the question, using it to plan a potential solution trajectory. You must conduct reasoning inside <think> and </think> first every time you get new information. After reasoning, if you find you lack some knowledge, you can call a search engine by <search> *query* </search> and it will return the top searched results between <information> and </information>. You can search as many times as you want, but make sure to avoid duplicate searches. If you find no further external knowledge is needed, you can directly provide the answer inside <answer> and </answer>, without detailed illustrations. For example, a short answer can be <answer> **Beijing** </answer> or longer such as <answer> **Sydney and Athens** </answer>.

Question: {question}

Figure 5: Prompt template used for search agent training in ORBIT. The agent interleaves <think> reasoning </think> blocks with <search> *query* </search> calls; retrieved passages are returned as <information> documents </information>, and the trajectory terminates when the model emits a final <answer> **answer** </answer>.

</answer> tag, whose content is matched against the ground-truth answer using exact-match (EM) scoring to produce the GRPO reward signal.

C ORBIT-4B Training Hyperparameters

We train the ORBIT-4B search agent using the `verl-tool` framework (Jiang et al., 2025), which extends VeRL with multi-turn, tool-augmented RL training capabilities. Training uses GRPO (Shao et al., 2024) with a live DDGS-based web retriever (top-k=5, parallel fan-out across `google`, `brave`, `bing`, `wikipedia`, and `grokipedia` backends). Rollouts are executed asynchronously via vLLM (v1 engine), with FSDP parameter and optimizer offloading to support the 8,192 token context window on a single 4×H100 SXM5 node. Table 7 summarizes the full configuration.

Module	Hyper-parameter	Value
Data	Max observation (search result) length	1,024
	Max response length	8,192
	Max prompt length	2,048
	Max action length	2,048
	Retriever top-k	5 (using DDGS)
Actor	Training batch size	256
	Mini-batch size	32
	Learning rate	1×10^{-6}
	LR warmup steps	10
	KL coefficient (β)	0.0
	KL loss type	`low_var_kl`
	Entropy coefficient	0.0
	Parallelism strategy	FSDP
Rollout	Max turns or search actions	5
	Group size G (rollouts per sample)	8
	Temperature	1.0
	Top-p	1.0
	vLLM GPU memory utilization	0.6
	vLLM max model length	8,192
	Rollout mode	Async (vLLM v1)
Reward	Reward function	Exact Match (EM)
	Mask observations in loss	Yes

Table 7: Hyperparameters used to train ORBIT-4B and other search agents (Search-R1-4B and InfoSeeker-4B) with GRPO via the `verl-tool` framework (Jiang et al., 2025).

Methods	Single-Hop QA			Multi-Hop QA				Avg. 7
	NQ	TQA	PopQA	HQA	2Wiki	MSQ	Bamb	
Search Agents (Base: Qwen2.5-3B)								
Search-o1 (Li et al., 2025b)	23.8	47.2	26.2	22.1	21.8	5.4	32.0	25.5
Search-R1-Instruct (Jin et al., 2025)	39.7	56.5	39.1	33.1	31.0	12.4	23.2	33.6
Search-R1-Base (Jin et al., 2025)	42.1	58.3	41.3	29.7	27.4	6.6	12.8	31.2
ReSearch-Instruct (Chen et al., 2025)	36.5	57.1	39.5	35.1	27.2	9.5	26.6	33.1
ReSearch-Base (Chen et al., 2025)	42.7	59.7	43.0	30.5	27.2	7.4	12.8	31.9
Dr. Zero-3B (Yue et al., 2026)	39.7	57.2	43.1	29.8	29.1	9.1	20.0	32.6
ZeroSearch-Base (Sun et al., 2025a)	43.0	61.6	41.4	33.8	34.6	13.0	13.9	34.5
StepSearch-Base (Zheng et al., 2025a)	–	–	–	32.9	33.9	18.1	32.8	–
EXSEARCH-Base (Shi et al., 2025b)	36.8	–	–	42.2	37.2	13.8	–	–
O^2-Searcher (Mei et al., 2025)	44.4	59.7	42.9	38.8	37.4	16.0	34.4	39.1
AutoRefine-Instruct (Shi et al., 2025a)	43.6	59.7	44.7	40.4	38.0	16.9	33.6	39.6
AutoRefine-Base (Shi et al., 2025a)	46.7	62.0	45.0	40.5	39.3	15.7	34.4	40.5
InForage (Qian & Liu, 2025)	42.1	59.7	45.2	40.9	42.8	17.2	36.0	40.5
CriticSearch (Zhang et al., 2025)	–	–	–	41.4	40.9	18.0	36.8	–
SE-Search (Li et al., 2026b)	**47.5**	62.4	42.3	45.0	36.1	18.3	42.4	42.0
SAPO-3B-Base (Li et al., 2026a)	47.4	63.0	44.9	44.9	41.2	19.6	42.4	43.3
SAPO-3B-Instruct (Li et al., 2026a)	46.9	62.9	45.7	**45.6**	44.8	**20.3**	43.2	44.2
Search Agents (Base: Qwen3-4B)								
Search-R1-4B†	37.9	58.8	40.7	34.3	37.1	13.0	44.8	38.1
InfoSeeker-4B†	38.1	63.8	42.1	40.4	41.1	15.7	48.0	41.3
ORBIT-4B†	44.7	**66.3**	**47.1**	42.3	**45.9**	19.4	**48.8**	**44.9**

Table 8: Accuracy comparison of search agents $\leq$ 4B parameters on single-hop and multi-hop QA benchmarks using the **Search-R1** evaluation setup (Jin et al., 2025), with the **2018 Wikipedia dump** (Karpukhin et al., 2020), **E5-base-v2** retriever (Wang et al., 2024) serving as the retrieval engine retrieving **top-3** documents for every search query. Qwen2.5-3B results are taken from SAPO (Li et al., 2026a). (†) denotes search agents trained in our work using Qwen3-4B as the base model. Best scores are highlighted in **bold**.

D Extended Comparison using the Search-R1 Evaluation Setup

We extend the ORBIT-4B results with a broader set of search agent baselines, by evaluating search agents trained on Qwen2.5-3B and our own baselines on Qwen3-4B by reproducing the Search-R1 evaluation setup (Jin et al., 2025). We retrieve passages from the 2018 Wikipedia dump (Karpukhin et al., 2020) using the E5-base-v2 retriever (Wang et al., 2024), retrieving top-3 passages at each turn. Results for search agents trained using Qwen2.5-3B as the base model are taken directly from SAPO (Li et al., 2026a), that evaluated these search agents in the Search-R1 setting (Jin et al., 2025).

Extended Results. Table 8 provides the EM accuracy of search agents evaluated by reproducing the Search-R1 evaluation setup (Jin et al., 2025), i.e., using the 2018 Wikipedia dump, E5-base-v2 retriever providing top-3 documents at each turn. From the results, we observe that ORBIT-4B achieves 44.9 EM accuracy outperforming InfoSeeker-4B by 3.6 points and Search-R1-4B by 6.8 points, on average. In addition, ORBIT-4B outperforms existing search agents trained with Qwen2.5-3B (Base or Instruct), demonstrating that using the recent and newer Qwen3-4B as the base model can outperform search agents trained with extensive training regimes with Qwen2.5-3B.

E ORBIT Dataset Examples

We present one representative training example from each of the 15 knowledge domains in ORBIT (dataset example for the domain *TV Shows & Movies* is already shown in Table 2). Colored spans highlight the individual factual clues embedded in each question; the verification section confirms every clue with the key supporting fact (shown in blue).

Answer Type	Definition
Temporal	Calendar date, year, month, decade, or explicitly time-anchored period.
Numeric / Quantitative	Numerical value, measurement, ratio, duration, or rate, with or without units.
Geographic Location	Physical place, region, landmark, or spatially identifiable location on Earth.
Person	Named individual (historical, contemporary, or fictional).
Organization / Institution	Named group, body, office, society, standardization entity, or collective actor.
Event / Historical Phenomenon	Named occurrence, movement, crisis, or episode that happens in time.
Scientific / Technical Concept	Formally defined concept, theory, method, model, or phenomenon.
Named Artifact / System	Specific non-human entity such as a tool, system, standard, protocol, document, product, currency, or legal instrument.
Property / Relationship	Characteristic, state, quality, constraint, or relational descriptor of an entity.

Table 9: Definitions of answer types used for categorizing answers in the ORBIT dataset.

Table 12: One training example from each of the 14 additional knowledge domains in ORBIT. Colored spans mark the distinct factual clues embedded in the question; the verification rows confirm each clue with key supporting evidence (blue).

Domain: *(2) Science & Technology*

Question: What term describes the mathematical concept first systematically developed in a 1939 physics textbook by a Nobel laureate who independently predicted antimatter, where angled brackets denote dual vector spaces and vertical bars separate state vectors, revolutionizing the formalism of quantum measurement theory?
Answer: bra-ket notation

Verification / Summary of Supporting Facts:
✓ **Origin:** *Paul Dirac introduced bra-ket notation in his 1939 publication* A New Notation for Quantum Mechanics.
✓ **Nobel Laureate:** *Dirac independently predicted antimatter (positrons) through his equation, confirmed experimentally in 1932.*
✓ **Notation:** *Uses angled brackets ⟨ (bras) and vertical bars | (kets) to denote dual vector spaces and separate state vectors.*
✓ **Impact:** *The notation revolutionized quantum measurement theory by formalizing state vectors, operators, and inner products.*

Domain: *(3) Art*

Question: In what year did the Cologne-based New Objectivity painter, whose work was confiscated during the state-sponsored campaign against "degenerate" art and who created psychologically charged portraits of women while married to an avant-garde artist known for depictions of solitary male figures, pass away after surviving wartime persecution and the destruction of her studio?
Answer: 1970

Verification / Summary of Supporting Facts:
✓ **Painter:** *Marta Hegemann, Cologne-based New Objectivity painter associated with the Cologne Progressives.*
✓ **Persecution:** *Her work was confiscated as "degenerate art" during the Nazi campaign.*
✓ **Style:** *She created psychologically charged portraits of women.*
✓ **Spouse:** *Married to Anton Räderscheidt, known for depictions of solitary male figures.*
✓ **Wartime:** *Endured wartime persecution and the destruction of her studio during WWII; died in obscurity in 1970.*

Domain: *(4) History*

continued on next page...

Table 12 continued from previous page

Question: What was the total number of personnel demobilized following the largest naval uprising in February 1946 that was directly inspired by the trial of soldiers from a colonial-era army unit, which itself became a catalyst for the final phase of a major 20th-century liberation struggle?
Answer: 476

Verification / Summary of Supporting Facts:
✓ **Uprising:** *The Royal Indian Navy mutiny (Feb. 1946) was the largest naval uprising, involving over 20,000 sailors.*
✓ **Trigger:** *Directly inspired by the trials of Indian National Army (INA) soldiers, a colonial-era unit.*
✓ **Impact:** *Acted as a catalyst for India's independence struggle, leading to British withdrawal in 1947.*
✓ **Outcome:** *Following the mutiny, 476 sailors were dismissed and were not reinstated post-independence.*

Domain: *(5) Sports*

Question: In the Summer Olympics subject to a major geopolitical boycott, which men's track event awarded gold to an athlete from a participating European nation who had set a world record in a different middle-distance event earlier that same year at a competition in a founding NATO country, and whose silver medalist represented a Soviet-bloc nation that defied the boycott?
Answer: Men's 1500 metres

Verification / Summary of Supporting Facts:
✓ **Olympics:** *The 1980 Moscow Olympics were boycotted by the United States and 65+ nations due to the Soviet invasion of Afghanistan.*
✓ **Gold Medalist:** *Sebastian Coe of Great Britain won gold in the men's 1500m; Britain participated in the Games.*
✓ **World Record:** *Coe set the 800m world record (1:42.33) in Oslo, Norway (a founding NATO member), in July 1979.*
✓ **Silver Medalist:** *Jürgen Straub represented East Germany, a Soviet-bloc nation that participated in the 1980 Games.*

Domain: *(6) Music*

Question: During what year was the only presentation made for a specialized Grammy category honoring dance music productions, awarded to a female vocalist's anthem of resilience recorded at a studio previously used by a British rock band for their live album featuring a cover of "Twist and Shout," coinciding with the ceremony where a singer-songwriter won Album of the Year?
Answer: 1980

Verification / Summary of Supporting Facts:
✓ **Grammy Category:** *The Grammy Award for Best Disco Recording was presented only once, at the 1980 ceremony.*
✓ **Winner:** *Gloria Gaynor's* I Will Survive *won the award; the song became a global anthem of resilience.*
✓ **Studio:** *Recorded at Atlantic Studios, where The Beatles tracked their 1963* Twist and Shout *EP cover.*
✓ **Ceremony:** *Christopher Cross won Album of the Year at the same 1980 Grammy Awards ceremony.*

Domain: *(7) Video Games*

Question: What is the name of the summon creature associated with a mountain location in a 1997 PlayStation RPG, where this entity's Japanese name derives from a historical Korean military unit and appears in a spin-off title directed by the developer who later created a 2006 action-RPG featuring a protagonist with amnesia?
Answer: Hiryū

Verification / Summary of Supporting Facts:
✓ **Origin:** Final Fantasy Tactics *(1997) includes the summon* Hiryū *associated with Mount Bervenia.*
✓ **Name:** *Hiryū's Japanese name derives from the historical Korean Hwarang military unit.*
✓ **Spin-off:** *Hiryū appears in* Final Fantasy XII: Revenant Wings*, directed by Yasumi Matsuno.*
✓ **Director:** *Yasumi Matsuno later directed* Final Fantasy XII *(2006), featuring protagonist Vaan with fragmented memories.*

Domain: *(8) Geography*

Question: What is the name of the city designated as World Book Capital for 2023, which became the first in its geographical region to receive this honor, known for hosting an annual international book fair that attracts publishers from French-speaking countries, and whose selection was announced by UNESCO in a city that shares its name with a famous aviation pioneer?
Answer: Accra

Verification / Summary of Supporting Facts:

continued on next page...

Table 12 continued from previous page

✓ **Designation:** *Accra was designated as the UNESCO World Book Capital for 2023.*
✓ **Region:** *First city in West Africa (third in Sub-Saharan Africa) to receive the title.*
✓ **Book Fair:** *Accra hosts the Ghana International Book Fair, attracting publishers from French-speaking countries.*
✓ **Announcement:** *Selection announced by UNESCO in Paris, the city named after the aviation pioneer Alberto Santos-Dumont's adopted home.*

Domain: *(9) Politics*

Question: At which major political convention did a coalition encompassing both moderate and conservative factions experience a platform dispute over foreign policy morality, resulting in the narrow defeat (by 60 votes) of a challenger who had advocated for that plank, during the last election cycle where this broad-coalition party failed to retain the presidency?
Answer: 1976 Republican National Convention

Verification / Summary of Supporting Facts:
✓ **Factions:** *Conflict between moderate President Gerald Ford and conservative challenger Ronald Reagan at the 1976 RNC.*
✓ **Platform Dispute:** *Reagan proposed a "foreign policy morality" plank criticising Ford's détente policy.*
✓ **Vote:** *Reagan's morality plank was narrowly defeated by approximately 60 votes at the convention.*
✓ **Election Outcome:** *The 1976 cycle ended with the Republican Party losing the presidency to Jimmy Carter.*

Domain: *(10) Medicine*

Question: What is the specific term for the phenomenon where certain small molecules form colloidal aggregates in aqueous buffer, leading to false positive results in biochemical assays that measure activity through optical detection methods commonly employed in large-scale compound evaluation?
Answer: colloidal aggregation

Verification / Summary of Supporting Facts:
✓ **Mechanism:** *Small molecules form colloidal aggregates in aqueous buffer, sequestering assay proteins.*
✓ **Consequence:** *Aggregates produce false positive results in biochemical assays by non-specifically inhibiting enzymes.*
✓ **Detection:** *Commonly detected via optical methods (absorbance/turbidity) in high-throughput screening.*
✓ **Context:** *A major artifact in large-scale compound evaluation (e.g., drug discovery HTS campaigns).*

Domain: *(11) Finance*

Question: Which financial valuation component equals the difference between a stock's market price and its no-growth value, requires forecasting of sustainable competitive advantages, incorporates dividend payout ratios in its calculation, is inversely related to the cost of equity capital, and becomes negative when a firm's return on equity falls below its required return?
Answer: PVGO

Verification / Summary of Supporting Facts:
✓ **Definition:** *PVGO (Present Value of Growth Opportunities) = Market Price − No-Growth Value.*
✓ **Input:** *Requires forecasting sustainable competitive advantages to estimate future excess returns.*
✓ **Formula:** *Incorporates dividend payout ratios in the Gordon Growth Model decomposition.*
✓ **Relationship:** *Inversely related to the cost of equity capital; a higher discount rate compresses PVGO.*
✓ **Sign:** *PVGO turns negative when ROE < required return, signalling value-destroying reinvestment.*

Domain: *(12) Law*

Question: What is the name of the landmark US Supreme Court case that upheld a state's authority to sterilize individuals deemed "unfit," where the majority opinion was authored by a justice who served concurrently with and frequently dissented from another justice famous for his "clear and present danger" test?
Answer: *Buck v. Bell*

Verification / Summary of Supporting Facts:
✓ **Case:** Buck v. Bell *upheld Virginia's authority to sterilize Carrie Buck under the 1924 Eugenical Sterilization Act.*
✓ **Author:** *Majority written by Justice Oliver Wendell Holmes Jr., declared "Three generations of imbeciles are enough."*
✓ **Colleague:** *Holmes served alongside Justice Louis Brandeis, co-author of the "clear and present danger" test (*Schenck v. United States, *1919).*
✓ **Legacy:** *The ruling remains unoverturned, enabling over 60,000 forced sterilizations across the United States.*

continued on next page...

Table 12 continued from previous page

Domain: *(13) Mathematics*

Question: What is the name of the measure of statistical dispersion often used for economic inequality that can be calculated using a method based on ranking and indexing, developed by an Italian statistician who also made significant contributions to the study of variability and demographic analysis?
Answer: Gini coefficient

Verification / Summary of Supporting Facts:
✓ **Measure:** *The Gini coefficient is the standard measure of statistical dispersion for economic inequality.*
✓ **Calculation:** *Computed via the Lorenz curve, which involves ranking income or wealth from lowest to highest.*
✓ **Author:** *Developed by Corrado Gini (1884–1965), Italian statistician.*
✓ **Contributions:** *Gini made foundational contributions to variability theory and demography.*

Domain: *(14) Puzzles*

Question: What is the title of the 1995 anti-gravity racing game developed and published by the same studio that released a puzzle game featuring suicidal green-haired creatures, which became a European launch title for a 32-bit console and featured licensed tracks from electronic musicians including a duo known for a tree-themed debut album?
Answer: Wipeout

Verification / Summary of Supporting Facts:
✓ **Game:** *Wipeout (1995) was an anti-gravity racing game developed by Psygnosis.*
✓ **Studio:** *Psygnosis also published* Lemmings *(1991), the puzzle game featuring green-haired suicidal creatures.*
✓ **Console:** *Wipeout was a European launch title for the PlayStation (32-bit) in 1995.*
✓ **Soundtrack:** *Featured licensed electronic tracks from Orbital and The Chemical Brothers.*

Domain: *(15) Code*

Question: What is the name of the algebraic structure formed by the equivalence classes of sentences under the relation that holds when two sentences have the same truth value in all possible interpretations, as introduced by a logician who also formulated a semantic definition of truth for formal languages and collaborated on a paradox involving the decomposition of spheres?
Answer: Lindenbaum algebra

Verification / Summary of Supporting Facts:
✓ **Structure:** *The Lindenbaum algebra consists of equivalence classes of sentences under logical equivalence.*
✓ **Equivalence:** *The relation holds when two sentences share the same truth value in all possible interpretations.*
✓ **Author:** *Alfred Tarski formulated the semantic definition of truth for formal languages and co-named the structure.*
✓ **Collaboration:** *Tarski collaborated with Stefan Banach on the Banach–Tarski paradox (decomposition of a sphere into two identical copies).*

Question–Answer Pair Generation Prompt

SYSTEM: You are an expert multi-hop factoid question creator. You should create complex or inverted questions containing answers that are easy to verify, but hard to solve for a given seed, along with the answer, a verification checklist, and a list of evidence URLs sufficient to justify the answer.

TASK: The question-answer pair generated should satisfy the conditions below:

- The question should have a unique answer.
- The answer should be a short fact (a term or a short phrase).
- The question should be very difficult to answer without the list of correct evidence URLs, requiring multiple web searches.
- The model answer should be easily verifiable using the evidence URLs and should appear as a substring in the final evidence.
- The question should be phrased so that each evidence does not give away too much information about the model answer so that the student can determine the model answer without finding all the evidences you've listed. That is, each evidence should be absolutely necessary to answer the question correctly.
- The question should be deep enough that it requires at least 5 evidence to answer the question correctly.
- You should not mention any proper nouns in the question as it gives away too much information about the model answer. Instead, describe the proper noun's properties in a way that is still enough to uniquely identify the proper noun.

PROCEDURE: You would typically start with a *seed* (could be a person, event, or artifact) that the user will provide. Find several characteristics within a large search space using the *seed*, and create a question from them.

REQUIRED OUTPUT: Please make sure you look into the seed carefully and think first before you provide your output in XML format. Make sure the *seed* is actually hidden in the question to make it challenging and your question should not be too verbose or easy to answer. Follow the format below by providing a verfied answer to the inverted question, a verification checklist containing citations containing evidence for each verification, and a list of evidence URLs:

```
<output>
    <inverted_question>your inverted query</inverted_question>
    <answer>the answer to the inverted query</answer>
    <verification_checklist>
        <item>bullet point 1 :cite[1]</item>
        <item>bullet point 2 :cite[1]</item>
        <item>bullet point 3 :cite[7]</item>
        etc.
    </verification_checklist>
    <evidence_urls>[1]: https://xxxx..., [2]: https://yyy..., [7]: https://zzzz...</evidence_urls>
</output>
```

EXAMPLE: `{Add an example here}`

NOTE: Make sure the *seed* is actually hidden in the question to make it challenging and also do not use the *seed* as your answer! Further, to ensure that your question is difficult enough, you should try to answer the question yourself from scratch, and see if you can answer it confidently by using less than 5 searches. If so, revise your question and try again.

SEED: `{seed}`

Figure 6: Prompt template for question–answer pair generation for a given input seed as inspiration and shuffled exemplars.

Self-Verification Prompt

QUESTION: `{question}`
ANSWER: `{answer}`

Given the question and answer, conduct a full web and Wikipedia search to retrieve the pages necessary to answer the question. Next, come up with a verification list of criteria, citing each URL for each criterion. Finally, provide a revised answer to the question (a short string) if needed, and lastly, list all the sources you cite with each URL and which all verification statement it supports!

Please do the web search first!

Figure 7: Prompt template for self-verification given the input question and answer.

(External-Verification) Answer Generation Prompt

SYSTEM: You are given a question and some evidence from scraped web documents. Please answer the question based on the evidence provided.

QUESTION: {question}

EVIDENCE: {URL: {url} Content: {web_content}, . . . }

Please provide your answer based on the evidence above.

(External-Verification) Answer Verification Prompt

SYSTEM: You are a judge evaluating whether a candidate response correctly answers a question, compared to a ground truth answer.

QUESTION: {question}

GROUND TRUTH ANSWER: {ground_truth_answer}

CANDIDATE RESPONSE: {predicted_answer}

Please judge whether the candidate response is correct or incorrect with respect to the ground truth answer. You should provide a brief explanation for your judgment, then output Judge: CORRECT or Judge: INCORRECT.

Figure 8: Prompt template for external verification in ORBIT, first the LLM independently generates an answer for the question given the evidence context. Next, the same LLM evaluates whether the predicted answer matches with the ground truth answer, and outputs either correct or incorrect.

Case Study on FRAMES (Krishna et al., 2025)

Question: What compass direction (of the 4 cardinal and 4 ordinal directions) is the capital city of the state which houses the Jackson Hole Mountain Resort in relation to the centre of the state?
Ground Truth: ["Southeast"]

Model Output: ORBIT-4B

<think> The question requires three reasoning hops: (1) identify the state that contains Jackson Hole Mountain Resort, (2) find the capital of that state, and (3) determine the capital's compass direction relative to the state's geographic centre. I recall that Jackson Hole lies in Wyoming; the capital is likely Cheyenne, but I should confirm. </think>

<search> *capital of Wyoming* </search>

<information> **Doc 1 (Title: "Wyoming State Capitol"):** The Wyoming State Capitol is located in the city of **Cheyenne**. Cheyenne is the **seat of government for the U.S. state of Wyoming**...
Doc 2 (Title: "Cheyenne, Wyoming"): Cheyenne ...is the **capital and most populous city of the U.S. state of Wyoming**, as well as the county seat of Laramie County, with 65,132 residents per the 2020 census...
Doc 3 (Title: "Wyoming"): The **state capital and most populous city is Cheyenne**... </information>

<think> Confirmed: the state is Wyoming and its capital is Cheyenne. I now need the geographic centre of Wyoming so I can determine the direction from that centre to Cheyenne. </think>

<search> *geographic center of Wyoming* </search>

<information> **Doc 3 (Title: "Outline of Wyoming"):** Geography of Wyoming — Wyoming is a U.S. state ...located in the Western United States ...
Doc 4 (Title: "Geography of Wyoming"): The Wind River Range in the **west central part** of the state...The Big Horn Mountains in the **north central portion**... </information>

<think> These results describe Wyoming's geography but do not give the geographic centre explicitly. I already know Wyoming is a roughly rectangular state; the centre is therefore approximately mid-state. Let me ask directly where Cheyenne sits relative to the rest of the state—that should encode the direction implicitly. </think>

<search> *direction of Cheyenne, Wyoming relative to center of state* </search>

<information> **Doc 1 (Title: "Cheyenne, Wyoming"): Lying near the southeast corner of the state**, Cheyenne is one of the least centrally located state capitals in the nation (together with cities such as Carson City, Nevada; Juneau, Alaska; Tallahassee, Florida; Topeka, Kansas; and Trenton, New Jersey).
Doc 3 (Title: "Laramie County, Wyoming"): The county seat is Cheyenne, the state capital. The county **lies west of the Nebraska state line and north of the Colorado state line** [i.e. in the southeast corner of Wyoming]. </information>

<think> The retrieved document explicitly states that Cheyenne lies near the *southeast corner* of Wyoming. The state's geographic centre therefore lies to the northwest of Cheyenne, which equivalently means Cheyenne is to the **southeast** of the centre. The answer is **southeast**. </think>

<answer> **Southeast** </answer>

Table 10: Case study of ORBIT-4B evaluated on FRAMES (Krishna et al., 2025). The search agent must chain geographic facts across multiple hops; key evidence extracted from retrieved documents is **bolded**.

Case Study on ORBIT during GRPO Training

Question: In which 1957 nature essay collection does a chapter describing a boy's unexpected discovery of a sunfish in a city creek—illustrating how preconceptions blind observers to novel phenomena—appear, written by an anthropologist-naturalist who later became president of the American Institute of Human Paleontology?
Ground Truth: ["The Immense Journey"]

Model Output: ORBIT-4B

<think> I need a 1957 nature essay collection with a sunfish chapter, written by an anthropologist-naturalist who later led the American Institute of Human Paleontology. This is a multi-hop question: (1) identify the author via the institute clue, (2) find the 1957 collection they wrote, and (3) confirm the sunfish chapter. Let me first search by the essay's contents. </think>

<search> *1957 nature essay collection sunfish city creek* </search>

<information> **Doc 1 (Title: "1957 — Wikipedia"):** 1957 (MCMLVII) was a common year starting on Tuesday of the Gregorian calendar…
Doc 2 (Title: "Historical Events in 1957"): Jan 22, 2013 — Historical events from year 1957. Learn about 514 famous, scandalous and important events…
Doc 5 (Title: "1957 Events & Facts"): MAJOR EVENTS: Soviet Union inaugurates the "Space Age" by launching Sputnik I… </information>

<think> Those results are not relevant—only generic 1957 almanac pages. Searching by essay content alone seems too obscure for the retriever. Let me try a more direct query about the specific episode. </think>

<search> *sunfish in a city creek essay 1957* </search>

<information> **Doc 1 (Title: "Sunfish Age and Hull Identification — SailingForums.com"):** …archives the full history of the Sunfish *sailboat* builders over the years…
Doc 3 (Title: "Sunfish vs Minifish? — SailingForums.com"): …The Minifish was easier for Skipper to scoot around in with higher winds vs wrestling the larger Sunfish sail… </information>

<think> The retriever conflates "sunfish" (a freshwater fish) with the *Sunfish* sailboat brand—both content-level searches have failed. I should change strategy and anchor on the author instead. The question provides a unique biographical clue: the author was president of the *American Institute of Human Paleontology*. Let me search for that role directly. </think>

<search> *anthropologist-naturalist who became president of American Institute of Human Paleontology* </search>

<information> **Doc 1 (Title: "Gigantopithecus — Wikipedia"):** Gigantopithecus has become popular in cryptozoology circles…
Doc 2 (Title: "The Immense Journey: An Imaginative Naturalist Explores the…"): Anthropologist and naturalist Loren Eiseley blends scientific knowledge and imaginative vision in this story of man. About the Author.
Doc 4 (Title: "Ernest Chantre: French anthropologist, naturalist and archaeologist"): Quick Facts …French anthropologist, naturalist and archaeologist. </information>

<think> Doc 2 is the breakthrough: it names **Loren Eiseley** as "anthropologist and naturalist" *and* mentions the title "**The Immense Journey**". I need to confirm this book was published in 1957. </think>

<search> *Loren Eiseley 1957 essay collection* </search>

<information> **Doc 4 (Title: "The Immense Journey by Loren Eiseley — The 5925th greatest book of…"):** This page contains details about the book **The Immense Journey** by **Loren Eiseley published in 1957**.
Doc 5 (Title: "The Immense Journey by Loren Eiseley (1957-08-12): Amazon.com…"): …Eiseley compares and comments on the growing understanding of evolution by the scientific community—**until 1957 when he wrote this**. </information>

<think> The evidence converges: *The Immense Journey* (1957), authored by anthropologist-naturalist *Loren Eiseley*, matches every clue in the question—the year, the genre (nature essays), and the author's disciplinary identity. The institutional presidency clue could not be directly confirmed from retrieved documents, but the triple match on year, genre, and author profile is conclusive. </think>
<answer> **The Immense Journey** </answer>

Table 11: Case study of ORBIT-4B during GRPO training on ORBIT. Key evidence extracted from retrieved documents is **bolded**.

Table 13: Top 100 categories per each domain in the ORBIT dataset generated using OpenAI's Deep Research tool.

List of up to 100 Wikipedia Categories (generated using OpenAI's Deep Research)
(1) TV Shows & Movies
American crime drama television series; British science fiction television series; Japanese drama television series; Anime television series; Sitcoms; Reality television series; Game shows; Miniseries; Documentary television series; Animated films; Action films; Comedy films; Horror films; Science fiction films; Documentary films; Silent films; Black-and-white films; Direct-to-video films; Short films; American independent films; Spaghetti Western films; Film noir; Kaiju; Blaxploitation films; Animated feature films; 1970s American television series; Hindi-language films; French New Wave; South Korean films; Musical films; Films about outer space; Films about immigration to the United States; Films about poverty; Films about educators; Films about viral outbreaks; Films about artificial intelligence; Films about time travel; Films based on actual events; Lost films; Unfinished films; Cancelled films; Directorial debut films; Film series; Film remakes; Hong Kong action films; Direct-to-video horror films; British comedy films; Indian horror films; French animated films; Anime films; Soap operas; Telenovelas; Netflix original programming; HBO original programming; Anthology television series; Serial killer films; Superhero films; Monster movies; Spy films; Biographical films; Fantasy films; War films; Disaster films; Adventure films; Thriller films; Psychological thriller films; Slasher films; Dystopian films; Post-apocalyptic films; Biblical films; Christmas films; Best Picture Academy Award winners; Film festivals; Television spin-offs; Unaired television pilots; Propaganda films; Crossover films; Video nasties; 3D films; IMAX films; Peplum films; Neo-noir films; Samurai films; Martial arts films; African films; Mexican films; Iranian films; Films based on video games; Television series based on films; Propaganda films; Crossover films; Film posters by genre; Remake television series; Unaired television pilots; Video nasties; 3D films; IMAX films.
(2) Science & Technology
Physics; Chemistry; Biology; Mathematics; Astronomy; Geology; Computer science; Engineering; Materials science; Nanotechnology; Robotics; Artificial intelligence; Machine learning; Quantum mechanics; Quantum field theory; General relativity; Thermodynamics; Particle physics; Nuclear physics; Optics; Astrophysics; Spaceflight; Satellites; Rocketry; Renewable energy; Climate change; Meteorology; Oceanography; Volcanology; Ecology; Evolutionary biology; Genetics; Bioinformatics; Neuroscience; Molecular biology; Virology; Biomedical engineering; Biotechnology; Genetic engineering; Cryonics; Telescopes; Microscopy; Chemical elements; Organic compounds; Polymers; Crystallography; Electronics; Semiconductors; Internet; Computer security; Operating systems; Programming languages; Open-source software; Computer networks; Supercomputers; Distributed computing; Cloud computing; Databases; Algorithms; Graph theory; Number theory; Combinatorics; Topology; Algebra; Geometry; Differential equations; Computational biology; Astrobiology; Exoplanets; Dark matter; String theory; Complex systems; Chaos theory; Computer vision; Natural language processing; Human–computer interaction; Embedded systems; Quantum computing; Virtual reality; Augmented reality; Blockchains; Mobile phones; Telecommunications; Software engineering; Programming paradigms; Data structures; Artificial life; Cryptography; Information theory; Game theory; Automation; Aerospace engineering; Mechanical engineering; Electrical engineering; Civil engineering; Chemical engineering; Computational chemistry; Systems biology; Computational complexity theory.
(3) Art
Continued on next page

List of up to 100 Wikipedia Categories (continued)

Painting; Sculpture; Photography; Architecture; Drawing; Printmaking; Ceramics; Textile arts; Glass art; Graphic design; Industrial design; Fashion design; Calligraphy; Comics; Theatre; Opera; Dance; Ballet; Circus; Magic (illusion); Performance art; Street art; Public art; Digital art; Contemporary art; Modern art; Renaissance art; Medieval art; Baroque art; Impressionism; Post-Impressionism; Surrealism; Cubism; Futurism; Dada; Abstract art; Abstract expressionism; Pop art; Art Nouveau; Gothic architecture; Baroque architecture; Modernist architecture; Skyscrapers; Video art; Installation art; Conceptual art; Murals; Portrait painting; Landscape art; Still life paintings; Mosaics; Poetry; Novels; Short stories; Plays; Science fiction novels; Fantasy novels; Detective novels; Children's literature; African art; Islamic art; Pre-Columbian art; Chinese art; Indian art; Japanese art; Indigenous Australian art; Mexican art; Native American art; Art history; Byzantine art; Christian art; Stagecraft; Choreography; Typography; Heraldry; Vexillology; Cartography; History painting; Nude art; Religious art; Decorative arts; Folk art; Outsider art; Manga; Musicals; Costume design; Black-and-white photography; Portrait photography; Renaissance architecture; Neoclassical architecture; Landscape architecture; French art; American art; Latin American art; Mime; Puppetry; Vaudeville; Improvisational theatre.

(4) History

Prehistory; Ancient history; Classical antiquity; Middle Ages; Renaissance; Age of Enlightenment; Age of Discovery; Industrial Revolution; French Revolution; American Revolutionary War; Napoleonic Wars; World War I; World War II; Cold War; Byzantine Empire; Ottoman Empire; Roman Empire; Ancient Egypt; Ancient Greece; Mesopotamia; Ming dynasty; Victorian era; Meiji period; Maya civilization; History of Europe; History of Asia; History of Africa; History of North America; History of the United States; History of China; History of India; History of science; History of medicine; Economic history; Military history; History of technology; History of literature; Maritime history; Historiography; Archaeology; Revolutions; Rebellions; Colonialism; Imperialism; Slavery; Crusades; European colonization of the Americas; Scramble for Africa; Exploration; Silk Road; Atlantic slave trade; Battles; Sieges; 19th century; 20th century; Bronze Age; Iron Age; Neolithic; Scientific Revolution; Space Age; Information Age; Space Race; Human migrations; Paleontology; World Heritage Sites; American Civil War; Russian Revolution; Spanish Civil War; The Holocaust; Genocide; Civil rights movement; Women's suffrage; Decolonization; Great Depression; Famines; Plagues; Age of Discovery; Achaemenid Empire; Mongol Empire; British Empire; Holy Roman Empire; Soviet Union; Cultural Revolution; Apartheid; American Old West; Sengoku period; History of printing; Illuminated manuscripts; Feudalism; History of New York City; History of the Catholic Church; History of Islam; History of Christianity; History of science fiction; History of mathematics; History of computing; Ancient warfare; Nuclear warfare.

(5) Sports

Association football; Basketball; Cricket; Baseball; American football; Rugby union; Rugby league; Tennis; Golf; Athletics (sport); Gymnastics; Swimming; Skiing; Ice hockey; Figure skating; Boxing; Martial arts; Mixed martial arts; Wrestling; Weightlifting; Cycling; Motorsport; Formula One; Chess; Extreme sports; Team sports; Individual sports; Combat sports; Water sports; Winter sports; Parasports; Women's sports; Youth sport; Olympic Games; Paralympic Games; Commonwealth Games; FIFA World Cup; UEFA European Championship; UEFA Champions League; Cricket World Cup; Super Bowl; World Series; Sport in India; Sport in the United States; College sports; High school sports; Olympic sports; Sports equipment; Sports venues; Doping in sports; Football hooliganism; Match fixing; Sports injuries; Mountaineering; Rock climbing; Underwater sports; Surfing; Skateboarding; Snowboarding; Shooting sports; Archery; Equestrian sports; Horse racing; Sumo; Kabaddi; Sport in Australia; International sports competitions; World championships; Motorcycle racing; Endurance sports; Marathons; Triathlon; Parkour; Professional wrestling; Orienteering; Table tennis; Badminton; Volleyball; Handball; Cue sports; Judo; Asian Games; Pan American Games; X Games; Winter Olympic Games; Summer Olympic Games; African Games; Speed skating; Karate; Taekwondo; Kickboxing; Esports.

(6) Music

Continued on next page

List of up to 100 Wikipedia Categories (continued)

Classical music; Rock music; Pop music; Hip hop music; Jazz; Blues; Country music; Electronic music; Heavy metal music; Punk rock; Folk music; Reggae; Soul music; Gospel music; Rhythm and blues; Funk; Latin music; K-pop; Hindustani music; Carnatic music; Japanese music; Brazilian music; Nigerian music; Baroque music; Classical period (music); Romantic music; 20th-century classical music; Contemporary classical music; Medieval music; Renaissance music; World music; Concept albums; Live albums; Soundtrack albums; Compilation albums; Debut albums; Instrumental music; A cappella; Choral music; Symphonies; Music theory; Musicology; Musical instruments; String instruments; Wind instruments; Percussion instruments; Keyboard instruments; Electronic musical instruments; Music festivals; Eurovision Song Contest; Music videos; Film scores; House music; Techno; Ambient music; Trance music; Drum and bass; Heavy metal genres; Progressive rock; Alternative rock; Hardcore punk; Doom metal; Death metal; Black metal; Thrash metal; Industrial music; New wave music; Disco; Salsa music; Tango music; Flamenco; Bhangra (music); Afrobeat; Highlife; J-pop; Mandopop; Barbershop music; Contemporary Christian music; Oratorios; Concertos; Operettas; National anthems; March music; Video game music; Chiptune; Music industry; Sound recording; Audio engineering; DJing; Hip hop genres; Trap music; Reggaeton; Karaoke; One-hit wonders; Christmas music; Patriotic songs; Protest songs; Ballads; Lullabies; Remix albums.

(7) Video Games

Action games; Adventure games; Role-playing video games; Simulation video games; Strategy video games; Puzzle video games; Platform games; Racing video games; Sports video games; First-person shooters; Third-person shooters; Survival horror video games; Stealth video games; Fighting games; Beat 'em up games; Visual novels; Massively multiplayer online games; Mobile games; Browser games; Arcade games; Indie video games; Open-source video games; Freeware games; Nintendo Switch games; PlayStation 4 games; Japanese video games; American video games; Indian video games; 2020 video games; Video games based on films; Video games based on novels; World War II video games; Educational video games; Christian video games; LGBTQ-related video games; Horror video games; Science fiction video games; Fantasy video games; Cyberpunk video games; Dystopian video games; Post-apocalyptic video games; Alternate history video games; Crossover video games; Game engines; Game design; Video game franchises; Mario video games; Pokémon video games; Sonic the Hedgehog video games; Final Fantasy video games; Resident Evil video games; Call of Duty video games; The Sims games; Tomb Raider games; Street Fighter games; Mortal Kombat video games; The Legend of Zelda games; Grand Theft Auto games; Star Wars video games; Star Trek games; DC Comics video games; Marvel Comics video games; Cross-platform video games; Roguelike video games; Metroidvania games; 4X video games; Tower defense video games; Dating sims; Incremental games; Virtual reality games; Augmented reality games; Text adventure games; Interactive fiction; Windows games; MacOS games; Linux games; Game Boy games; PlayStation 2 games; Xbox One games; Video game controversies; Video games based on Marvel Comics; Video games based on DC Comics; Esports games; Cancelled video games; Dōjin games; Erotic video games; Multiplayer online battle arena games; Battle royale games; Open world video games; Action-adventure games; Shoot 'em up games; Rhythm games; Music video games; Video game compilations; Game jam video games; Video game mods; Unreal Engine games; Unity (game engine) games.

(8) Geography

Geography of Africa; Geography of Asia; Geography of Europe; Geography of North America; Geography of South America; Geography of Oceania; Geography of Antarctica; Mountains; Rivers; Lakes; Islands; Deserts; Forests; Volcanoes; Caves; Canyons; Waterfalls; Glaciers; Oceans; Seas; National parks; Protected areas; Ghost towns; Megacities; Capital cities; Ports and harbours; Bridges; Tunnels; Mountain ranges; Seamounts; Oceanic trenches; Atolls; Coral reefs; Geographic poles; Extreme points of Earth; Maps; Earthquakes; Tsunamis; Tropical cyclones; Volcanic eruptions; Impact craters; Middle East; Scandinavia; Latin America; Arctic; Sahara; Himalayas; Alps; Andes; Amazon basin; Great Plains; Siberia; Pacific Ocean; Atlantic Ocean; Indian Ocean; Arctic Ocean; Southern Ocean; Mediterranean Sea; Caribbean Sea; Red Sea; Bays; Peninsulas; Valleys; Deltas; Swamps; Steppes; Plateaus; Oases; Ramsar sites; Urban planning; Time zones; Equator; Tropics; Reservoirs; Dams; Straits; Isthmuses; Beaches; Sand dunes; Mountain passes; Geysers; Countries in Africa; States of the United States; Metropolitan areas; Volcanic islands; Highlands; Rainforests; Tundra; Savannas; Monsoons; Climate zones; Permafrost; Settlements; Villages; Towns; Continents; Borders; Coasts; Transcontinental countries; Natural disasters.

(9) Politics

Continued on next page

List of up to 100 Wikipedia Categories (continued)

Communism; Socialism; Capitalism; Anarchism; Conservatism; Liberalism; Fascism; Democracy; Monarchy; Republics; Parliamentary systems; Totalitarianism; Populism; Nationalism; Colonialism; Imperialism; Globalization; Human rights; Constitutions; Legislatures; Elections; Referendums; Political parties; Political ideologies; Revolutions; Coups d'état; Propaganda; Censorship; Political scandals; Corruption; Authoritarianism; Militarism; Separatism; Federalism; Unitary states; Dictatorships; United Nations; NATO; Non-governmental organizations; International relations; Diplomacy; Treaties; War; Peace movements; Disarmament; Genocides; Electoral systems; Public policy; Constitutional courts; Political campaigns; Lobbying; Political violence; Terrorism; Ideological conflicts; Civil wars; Social movements; Civil rights; Women's rights; LGBT rights; Environmentalism; Climate politics; Immigration; Gun politics; Drug policy; Healthcare reform; Education policy; Tax policy; City politics; Local government; Federal governments; Politics of the United States; Politics of the United Kingdom; Politics of India; Politics of France; Politics of the European Union; Politics of the United Nations; Elections by country; Political parties by country; Revolutions by country; Human rights by country; Military coups by country; Political corruption by country; Foreign relations by country; Populism by country; Nationalism by country; Cold War; War on Terror; Arab Spring; Democratization; Electoral fraud; Gerrymandering; Propaganda by country; Public opinion; Political communication; Freedom of speech; Campaign finance.

(10) Medicine

Diseases; Infectious diseases; Genetic disorders; Cardiovascular diseases; Cancers; Neurological disorders; Mental disorders; Respiratory diseases; Skin conditions; Pandemics; Rare diseases; Pediatrics; Geriatrics; Oncology; Cardiology; Neurology; Psychiatry; Endocrinology; Immunology; Epidemiology; Public health; Preventive medicine; Occupational safety and health; Pharmacology; Clinical trials; Vaccines; Antibiotics; Medical devices; Prosthetics; Medical imaging; Diagnostics; First aid; Emergency medicine; Surgery; Anesthesiology; Transplantation medicine; Medical genetics; Pathology; Radiology; Dermatology; Ophthalmology; Obstetrics and gynecology; Orthopedics; Sports medicine; Dentistry; Veterinary medicine; Alternative medicine; Traditional medicine; Herbalism; Homeopathy; Chiropractic; Medical ethics; History of medicine; Medical education; Hospitals; Healthcare in the United States; Healthcare in the United Kingdom; Healthcare in India; Health law; Medical research; Evidence-based medicine; Global health; Nutrition; Occupational therapy; Physical therapy; Rehabilitation medicine; Pharmaceuticals; Drug discovery; Medical imaging; Telemedicine; Genomics; Stem cell research; Bioethics; Clinical psychology; Disease outbreaks; Quarantine; Tropical diseases; Neglected diseases; Zoonotic diseases; Endemic diseases; Chronic diseases; Terminal illnesses; Palliative care; Mortality; Surgery by specialty; Minimally invasive surgery; Robotic surgery; Plastic surgery; Medical imaging; Radiotherapy; Chemotherapy; Pharmacology; Over-the-counter drugs; Prescription drugs; Controlled substances; Addiction medicine; Outbreaks.

(11) Finance

Finance; Economics; Macroeconomics; Microeconomics; Banking; Accounting; Personal finance; Corporate finance; Public finance; International finance; Financial markets; Stock market; Bond market; Foreign exchange market; Derivatives (finance); Commodities; Cryptocurrencies; Insurance; Real estate; Investing; Investment funds; Hedge funds; Private equity; Venture capital; Mutual funds; Index funds; Exchange-traded funds; Retirement plans; Pension funds; Wealth management; Asset management; Risk management; Financial risk; Credit; Loans; Mortgages; Interest rates; Inflation; Deflation; Central banks; Monetary policy; Fiscal policy; Economic growth; Recessions; Financial crises; Stock market crashes; Depressions (economics); Economic bubbles; Hyperinflation; International trade; Balance of payments; Exchange rates; Budget deficits; Public debt; Taxation; Tariffs; Economic history; Financial history; Economic systems; Capitalism; Socialism; Mixed economies; Market economies; Planned economies; Shadow economy; Underground economy; Financial regulation; Financial law; Bank regulation; Anti-money laundering; Financial technology; Digital currencies; Mobile payments; Online banking; High-frequency trading; Stock exchanges; Derivatives trading; Commodity markets; Real estate investment trusts; Housing market; Insurance companies; Reinsurance; Actuarial science; Corporate governance; Shareholder activism; Financial scandals; Insider trading; Accounting scandals; Pyramid schemes; Fraud; Ponzi schemes; Financial literacy; Personal budgeting; Philanthropy; Microfinance

(12) Law

Continued on next page

List of up to 100 Wikipedia Categories (continued)

Criminal law; Civil law (legal system); Common law; Constitutional law; Administrative law; International law; Human rights law; Labour law; Environmental law; Intellectual property law; Patent law; Copyright law; Trademark law; Contract law; Tort law; Property law; Family law; Inheritance law; Tax law; Bankruptcy law; Immigration law; Criminal justice; Law enforcement; Evidence law; Legal history; Trials; Case law; Law of the United States; Law of the United Kingdom; Law of India; Sharia; Canon law; Roman law; Jurisprudence; Legal doctrines; Legal ethics; Law reform; War crimes; International criminal law; International humanitarian law; Crimes; Punishments; Prisons; Execution methods; Capital punishment; European Court of Human Rights cases; United States Supreme Court cases; International Court of Justice cases; Landmark court decisions; Arbitration; Mediation; Lawsuits; Class action lawsuits; Corruption; Money laundering; Counterterrorism; Cybercrime; Organized crime; Criminology; Forensic science; Fraud; Gun laws; Censorship; Drug control law; Martial law; Anti-discrimination law; Privacy law; Internet law; Space law; Law of the sea; Banking law; Insurance law; Competition law; Commercial law; Admiralty law; Election law; Natural law; Military law; Obscenity law; Same-sex marriage law; Abortion law; Blue laws; Legal codes; Constitutions; Supreme courts; Courts by country; Police misconduct; Wrongful convictions; Freedom of expression law; Blasphemy law; Indigenous law; Religious law; Customary law; Legal fictions; Extradition; International criminal tribunals; Arms control; International sanctions; Geneva Conventions; Nuremberg trials

(13) Mathematics

Algebra; Mathematical analysis; Applied mathematics; Arithmetic; Calculus; Combinatorics; Computational mathematics; Discrete mathematics; Dynamical systems; Elementary mathematics; Experimental mathematics; Foundations of mathematics; Game theory; Geometry; Graph theory; Mathematical logic; Number theory; Order theory; Recreational mathematics; Topology; Probability; Statistics; Abstract algebra; Linear algebra; Prime numbers; Euclidean geometry; Non-Euclidean geometry; Algebraic geometry; Differential geometry; Algebraic topology; Mathematical physics; Chaos theory; Set theory; Category theory; Differential equations; Partial differential equations; History of mathematics; Philosophy of mathematics; Mathematics education; Women in mathematics; Mathematicians; Mathematics competitions; Mathematics organizations; Mathematics awards; Mathematical games; Theoretical computer science; Information theory; Fractals; Numbers; Probability distributions; Mathematical notation; Mathematical theorems; Mathematical proofs; Pseudomathematics; Mathematical tools; Mathematics and art; Mathematics and culture; Mathematical constants; Mathematical concepts; Polygons; Polyhedra; Curves; Surfaces; Polynomials; Special functions; Equations; Inequalities (mathematics); Temporal paradoxes; Philosophical paradoxes; Undecidable problems; NP-complete problems; Computational complexity theory; Combinatorial optimization; Linear programming; Coding theory; Cryptographic algorithms; Cryptographic attacks; Ciphers; Classical cryptography; Decipherment; Disk encryption; Cryptographic hardware; History of cryptography; Key management; Cryptography law; National Security Agency cryptography; Cryptography organizations; Cryptographic primitives; Cryptographic protocols; Public-key cryptography; Quantum cryptography; Security engineering; Cryptographic software; Cryptography standards; Steganography; Symmetric-key cryptography; Theory of cryptography.

(14) Puzzles

Adventure game puzzles; Anagrams; Puzzle books; Chess problems; Puzzle competitions; Cryptography contests; Puzzle designers; Escape rooms; Guessing games; Locked-room mysteries; Puzzle magazines; Puzzle manufacturers; Mathematical puzzles; Mazes; Mechanical puzzles; Paradoxes; Riddles; Tiling puzzles; Unsolvable puzzles; Puzzle video games; Word puzzles; Logic puzzles; Nonograms; Rubik's Cube; Shunting puzzles; Sudoku; Combination puzzles; Jigsaw puzzles; Mechanical puzzle cubes; Chess composers; Fairy chess; Mathematical chess problems; Folklore featuring impossible tasks; Koan; Puzzle hunts; Alternate reality games; The Amazing Race; Crosswords; Word puzzle video games; Puzzle video games by series; Logical paradoxes; Mathematical paradoxes; Optical illusions; Hidden object games; Tile-matching video games; Sudoku video games; Maze games; Falling block puzzle games; Mahjong video games; Professor Layton; Puyo Puyo; Tetris; Tetris games; Puzzle Bobble; Bejeweled; Angry Birds; Portal (series); Myst games; Bomberman; Lemmings games; Brain teasers; Illusions; Auditory illusions; Tactile illusions; Dilemmas.

(15) Code

Continued on next page

List of up to 100 Wikipedia Categories (continued)

Algorithms; Anti-patterns; Programming contests; Code refactoring; Concurrent computing; Programming constructs; Program derivation; Data structures; Debugging; Computer programming folklore; Programming games; Programming idioms; Programming language syntax; Programming language topics; Programming languages; Computer libraries; Live coding; Programming paradigms; Programming principles; Software optimization; Quantum programming; Self-hosting software; Software design; Software design patterns; Source code; Source code generation; Streaming algorithms; Programming tools; Visual programming; Computer programmers; Computer programming books; Computer programming stubs; Cryptography; Cryptography lists and comparisons; Cryptographic algorithms; Applications of cryptography; Cryptographic attacks; Ciphers; Classical cryptography; Cryptographers; Decipherment; Disk encryption; Cryptologic education; End-to-end encryption; Cryptographic hardware; History of cryptography; Key management; Cryptography law; National Security Agency cryptography; Cryptography organizations; Cryptographic primitives; Cryptographic protocols; Public-key cryptography; Cryptography publications; Quantum cryptography; Security engineering; Cryptographic software; Cryptography standards; Steganography; Symmetric-key cryptography; Theory of cryptography; Cryptography stubs; Computer programmers; Algorithms; Anti-patterns; Programming contests; Code refactoring; Concurrent computing; Programming constructs; Program derivation; Data structures; Debugging; DLL injection; Computer programming folklore; Programming games; Programming idioms; Programming interfaces; Programming language syntax; Programming language topics; Programming languages; Computer libraries; Live coding; Method (computer programming); Programming paradigms; Programming principles; Software optimization; Quantum programming; Self-hosting software; Software design; Software design patterns; Source code; Source code generation; Streaming algorithms; Programming tools; Visual programming.