

Think, Search, Correct: Adaptive Tool-Use Agents

Megha Kataki
mkataki@ucsd.edu

Arshia Vadhani
avadhani@ucsd.edu

Joyce Lu
jol072@ucsd.edu

Jiaxin Yang
jiy016@ucsd.edu

Kun Zhou
kuzhou@ucsd.edu

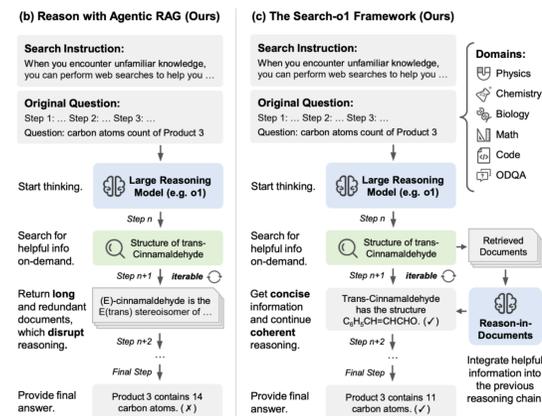
Domain Context

Domain Problem- Large Language Model (LLMs) agents have the ability to utilize tools via API calls to fetch external information, but most of them treat tool calls with a fixed pattern of logic resulting in inaccuracy and inefficiency

Our Solution- Update these search agents internally for better external multi-model steps. Implement a judge and reflection function at each step of iterative search to evaluate whether information extracted is true, relevant, and effectively found. With each step seek to identify issues with current search queries and prompt engineer the model to find better information to wholly answer the question.

Introduction

- Baseline Framework - Search-o1:** Unlike standard RAG (Retrieval-Augmented Generation), Search-o1 iteratively retrieves and analyzes information through its Reason-in-Documents module, updating its reasoning trajectory as new evidence is found.



- Problem Statement:** We conducted a case study on the Search-o1 architecture and identified three architectural weaknesses:
 - Hallucination: Model skipped necessary search calls, leading to unsupported final answers.
 - Extraction Information loss: Extraction module occasionally failed to preserve relevant evidence, causing information loss when useful documents were retrieved.
 - Ineffective Search Planning: Multi-hop search planning mechanism was ineffective in correcting earlier reasoning mistakes.

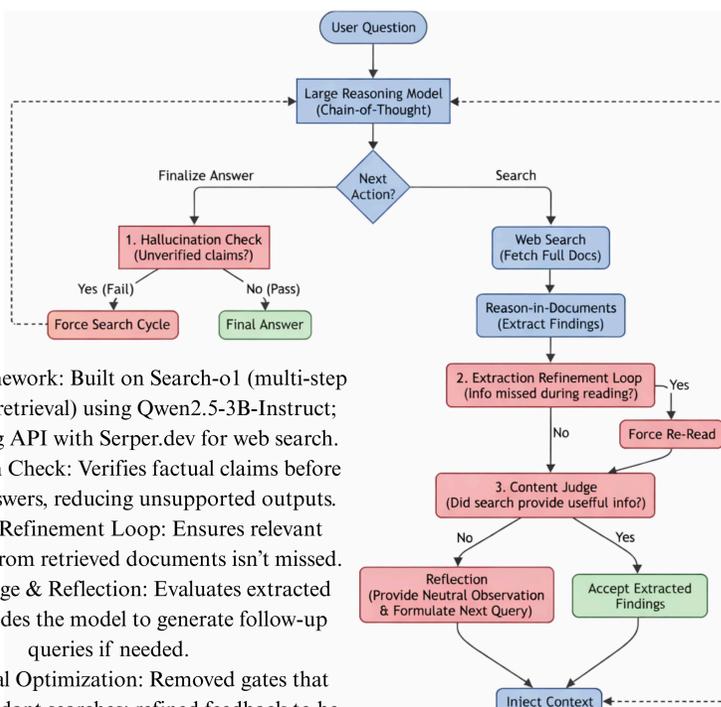
Data

HotpotQA (requires multi-hop reasoning)

Multi-hop: Questions that cannot be answered using a single source and instead require gathering and connecting information from multiple sources.

- Example: "What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell?"
- Infrastructure: Qwen2.5-3B-Instruct (Backbone LLM), Serper.dev (Search API), Jina API (Fetch full page content)

Methods



- Baseline Framework:** Built on Search-o1 (multi-step reasoning + retrieval) using Qwen2.5-3B-Instruct; replaced Bing API with Serper.dev for web search.
- Hallucination Check:** Verifies factual claims before finalizing answers, reducing unsupported outputs.
- Extraction Refinement Loop:** Ensures relevant information from retrieved documents isn't missed.
- Content Judge & Reflection:** Evaluates extracted info and guides the model to generate follow-up queries if needed.
- Architectural Optimization:** Removed gates that caused redundant searches; refined feedback to be constructive rather than punitive.
- Evaluation Setup:** Subset of HotpotQA (120 samples), measured Accuracy & F1; step-by-step reasoning and extraction history logged for analysis.

Qualitative Results

```
id 28
Question *cjm_start-system You are Owen, created by Alibaba Cloud. You are a helpful assistant.<end_cjm_end><lm_start-user You are a reasoning assistant with the ability to perform web searches to help you answer the user's question accurately. You have special tools: - To perform a search: write <begin_search_query> your query here <end_search_query>. Then, the system will search and analyze relevant web pages, then provide you with helpful information in the format <begin_search_result> ...search results...<end_search_result>. You can repeat the search process multiple times if necessary. The maximum number of search attempts is limited to 10. Once you have all the information you need, continue your reasoning. Example: Question: "Alice David is the voice of Lara Croft in a video game developed by which company?" Assistant thinking steps: - I need to find out who voices Lara Croft in the video game. - Then, I need to determine which company developed that video game. Assistant: <begin_search_query>Alice David Lara Croft voice<end_search_query> (System returns processed information from relevant web pages) Assistant thinks: The search results indicate that Alice David is the voice of Lara Croft in a specific video game. Now, I need to find out which company developed that game. Assistant: <begin_search_query>video game developed by Alice David Lara Croft<end_search_query> (System returns processed information from relevant web pages) Assistant continues reasoning with the new information. Remember: - Use <begin_search_query> to request a web search and end with <end_search_query>. - When done searching, continue your reasoning. Please answer the following question. You should think step by step to solve it. Provide your final answer in the format [boxed]YOUR ANSWER. Question: Kaiser Ventures corporation was founded by an American industrialist who became known as the father of modern American shipbuilding? <lm_end><lm_start>assistant
answer [ ] 1 item
Output *To answer this question, let's break it down into steps and search for relevant information: 1. **Identify the American industrialist who became known as the father of modern American shipbuilding.** - We need to find out who this person was. 2. **Determine if Kaiser Ventures corporation was founded by this industrialist.** - Once we identify the industrialist, we need to verify if Kaiser Ventures was indeed founded by them. Let's start with step 1 by searching for the person known as the father of modern American shipbuilding. <begin_search_query>father of modern American shipbuilding industrialist<end_search_query> <begin_search_result>Henry J. Kaiser, born on May 9, 1882, in Sprout Brook, New York, was an American industrialist who became known as the father of modern American shipbuilding. According to various sources, Kaiser played a significant role in the construction of numerous infrastructure projects, including roads, dams, and highways. His company, the Henry J. Kaiser Company, was established in 1914 and later grew to become a major player in the construction and shipbuilding industries. In 1931, Kaiser's company won a contract to build the Hoover Dam, demonstrating his capabilities in large-scale engineering projects. His company, Kaiser Shipyards, was instrumental in building Liberty ships during World War II, contributing significantly to the war effort. <end_search_result> <Observation: Useful information or candidates found. Proceed with verification or next step.) From the search results, it is clear that Henry J. Kaiser was the industrialist known as the father of modern American shipbuilding. Now, let's move to step 2 and verify if Kaiser Ventures corporation was founded by Henry J. Kaiser. <begin_search_query>Kaiser Ventures corporation founder Henry J. Kaiser<end_search_query> <begin_search_result>No helpful information found.<end_search_result> <Observation: The search results were limited. Suggestion: ## Analysis: The search failed because the terms used in the previous search are not directly aligned with the information we need. The query "Henry J. Kaiser founder Kaiser Ventures shipbuilding industry" might be too specific and might not yield the right results. Additionally, "Kaiser Ventures" might not be the primary focus, and "shipbuilding industry" alone might not pinpoint the right person. The user is asking about an American industrialist known as the "father of modern American shipbuilding" who founded Kaiser Ventures. This implies we need to focus on the industrialist, the shipbuilding aspect, and the founding of the corporation. However, the emphasis seems to be on identifying the industrialist first. ## Next Query: "Industrialist known as the father of American shipbuilding founder of Kaiser Ventures" This query is more generalized and aligns better with the user's question. It breaks down the main points: - An industrialist - Known as the father of American shipbuilding - Founder of Kaiser Ventures This should help in finding the relevant information more effectively.) <begin_search_query>Industrialist known as the father of American shipbuilding founder of Kaiser Ventures<end_search_query> <begin_search_result>Henry J. Kaiser is known as the father of modern American shipbuilding and played a significant role in the construction of numerous infrastructure projects, including roads, dams, and highways. His company, Kaiser Shipyards, built Liberty ships during World War II, contributing significantly to the war effort. <end_search_result> <Observation: Useful information or candidates found. Proceed with verification or next step.) Based on the search results, Henry J. Kaiser is indeed the industrialist known as the father of modern American shipbuilding and he founded Kaiser Shipyards, which built Liberty ships during World War II. Given that Kaiser Ventures corporation was established in 1931 and is known for its involvement in various ventures, it's plausible that Henry J. Kaiser was also involved in founding or overseeing the company. To conclude, Henry J. Kaiser, the industrialist known as the father of modern American shipbuilding, founded Kaiser Ventures corporation. Final Answer: [boxed]Henry J. Kaiser
```

Here we show an example multi-hop prompt along with the reasoning generated by our model. The output includes the model's reasoning process and the extracted information used to generate the final response. For each multi-hop query, we compare the expected answer with the predicted answer produced by the model. We also display the document extraction and reasoning steps, which illustrate how the model combines information from multiple sources to arrive at its prediction.

Note: Exact Match was omitted because the reasoning model may generate answers with different wording while still being correct. F1 score better captures this semantic overlap between predicted and expected answers.

Quantitative Results

We evaluated the model outputs using Accuracy and F1 score to capture both overall correctness and balance between precision and recall.

- Accuracy captures overall correctness.
- F1 ensures the model performs well on both common and rare outputs.

Evaluation Metric	Our Model Search-o1 Acc	Baseline Search-o1 Acc	Our Model Search-o1 F1	Baseline Search-o1 F1
Average	0.43	0.29	0.46	0.36

Conclusion

We improved current baselines by adding refinement checks during the retrieval stage, resulting in measurable gains in F1 score and accuracy. While AI agents are powerful tools, it is important to ensure they are functioning correctly.

Our evaluation shows that accuracy and F1 alone are insufficient for fully assessing performance; semantic evaluation is necessary to understand whether outputs truly capture intended meaning. The refinement step leverages a large reasoning model with prompt engineering, highlighting the importance of carefully designing prompts to optimize results.

For future work, we aim to:

- Explore different prompt engineering strategies.
- Incorporate explicit semantic evaluation metrics.
- Address remaining bottlenecks in retrieval and refinement to further improve performance.

Impact

Let's say you have a toolbox and you grab a screwdriver at random, but your goal is to hammer a nail. Even though you grabbed a tool, it wasn't pertinent to your current task. Through our work, we wanted to similarly demonstrate that using a tool is not the only necessary mechanism for achieving smarter results.

Including a verification step to improve accuracy and efficiency is pivotal when creating these types of models, and can help avoid hallucinations and unnecessary computations. For the future, this type of approach could help the reliability of AI systems that are used in real-world applications such as education, healthcare, and decision support.

References

- [1] Li, Xiaoxi, et al. "Search-o1: Agentic search-enhanced large reasoning models." Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025.
- [2] Yang, Zhilin, et al. "HotpotQA: A dataset for diverse, explainable multi-hop question answering." Proceedings of the 2018 conference on empirical methods in natural language processing. 2018.
- [3] Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W. W., Salakhutdinov, R., & Manning, C. D. (2018). HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. arXiv:1809.09600.

Website QR

