

# Agentic Reasoning: Reasoning LLMs with Tools for the Deep Research

Junde Wu, Jiayuan Zhu, Yuyuan Liu  
University of Oxford

## Abstract

In this technical report, we introduce Agentic Reasoning, a framework<sup>1</sup> that enhances large language model (LLM) reasoning by integrating external tool-using agents. Unlike conventional LLM-based reasoning approaches, which rely solely on internal inference, Agentic Reasoning dynamically engages web search, code execution, and structured reasoning-context memory to solve complex problems requiring deep research and multi-step logical deduction. Our framework introduces the Mind Map agent, which constructs a structured knowledge graph to track logical relationships, improving deductive reasoning.

Additionally, the integration of web-search and coding agents enables real-time retrieval and computational analysis, enhancing reasoning accuracy and decision-making. Evaluations on PhD-level scientific reasoning (GPQ) and domain-specific deep research tasks demonstrate that our approach significantly outperforms existing models, including leading retrieval-augmented generation (RAG) systems and closed-source LLMs. Moreover, our results indicate that agentic reasoning improves expert-level knowledge synthesis, test-time scalability, and structured problem-solving. The code is at: [https://github.com/theworldofagents/Agentic Reasoning](https://github.com/theworldofagents/Agentic_Reasoning).

## 1 Introduction

Recently, large reasoning models, such as OpenAI’s o1 (Jaech et al., 2024), Qwen-QwQ (Team, 2024), and DeepSeek-R1 (Team, 2024), have demonstrated impressive stepwise reasoning capabilities over long sequences through large-scale reinforcement learning. These advancements provide promising solutions to complex reasoning tasks (Wei et al., 2022; Lewkowycz et al., 2022; OpenAI) and have inspired foundational efforts to replicate

o1-like reasoning patterns across a broader range of models (Qin et al., 2024; Huang et al., 2024; Zhang et al., 2024).

DeepSeek-R1, for example, relies exclusively on rule-based outcome rewards during training, such as evaluating whether a mathematical solution is correct or a piece of code executes successfully. While this approach has yielded remarkable reasoning capabilities, equaling o1’s performance in domains like math and code, it comes with notable trade-offs. As even the authors acknowledge, this type of training diminishes the model’s ability to articulate its reasoning process. DeepSeek-R1’s responses are often logical and accurate but lack detailed explanations of transitions between ideas or the finer connections between arguments.

Although current reasoning methods excel in structured domains like math and code—where outcomes are easily verifiable—applying these techniques to less structured or subjective tasks remains a significant challenge. Adapting these strategies to areas where answers are not inherently definitive is a key research gap. How can models be trained to handle tasks that require judgment, interpretation, or nuanced understanding rather than binary correctness?

Furthermore, not all problems benefit from formal reasoning approaches. Many fields, such as social sciences, ethics, or experiential disciplines, rely on abstract concepts, conventional wisdom, factual verification, understanding complex logical relationships, or moral reasoning. When models attempt to impose math- or coding-style reasoning onto such areas, they often produce flawed or overly rigid results. Developing approaches that account for these unique requirements is essential for advancing the applicability of reasoning model beyond their current domains.

Deep, thoughtful answers to open-ended questions often require extensive research, repeated verification, information retrieval, computational anal-

<sup>1</sup>work in progress

ysis, and the organization of complex logical relationships—steps fundamental to human reasoning. In this process, humans rely heavily on external tools, such as internet searches for gathering information, computational tools for quantitative analysis, or whiteboards and Mind Maps for organizing thoughts. This raises an intriguing question: can large language models similarly leverage external tools to enhance their reasoning and tackle intensive knowledge work across diverse domains?

Previous efforts have attempted to integrate search or retrieval-augmented generation (R<sub>G</sub>) into the reasoning process (Shao et al., 2024; Khaliq et al., 2024; Islam et al., 2024; Li et al., 2025), with notable examples including Gemini’s Deep Research. However, these models are closed, their exact methodologies remain undisclosed. In contrast, open-source models typically focus exclusively on retrieval or web-searching during reasoning, leaving a significant performance gap compared to their closed-source counterparts.

We introduce *agentic Reasoning*, a framework that enhances the reasoning process by integrating external LLM-based agents as tools. This approach enables LLMs to perform multi-step reasoning and tackle complex problems more effectively by delegating specific tasks to these auxiliary agents. Through extensive experimentation with integrating various agents into the reasoning process, we identified three essential agents that prove highly effective for general reasoning across diverse problems. The web-search agent, which retrieves relevant information from the internet to supplement the model’s knowledge. The code agent, capable of performing computational analyses and coding tasks to support quantitative reasoning. Finally, the memory agent, which we call Mind Map, constructs knowledge graphs based on the reasoning context, enabling the organization of complex logical relationships in a manner similar to human mind mapping. Together, these agents enhance the model’s ability to tackle complex problems with greater efficiency and precision.

When integrated into current reasoning LLMs, *agentic Reasoning* transforms their problem-solving capabilities by enabling them to plan and execute multi-step strategies autonomously. These models can identify and retrieve the necessary data, adapt dynamically to real-time information, and perform quantitative analyses to generate precise outcomes. This framework also allows LLMs to deliver comprehensive reports comparable to those

of a research analyst or provide solutions on par with PhD-level expertise.

We evaluated our model on general knowledge-intensive benchmarks requiring complex reasoning capabilities, categorized into two key areas: (1) solving expert-level questions and (2) conducting deep research on real-world expert-level tasks.

For expert-level questions, we tested the model on the GPQ dataset, a PhD-level science multiple-choice Q<sub>A</sub> benchmark with questions authored by domain experts in physics, chemistry, and biology. Our *agentic Reasoning* framework achieved impressive accuracy rates: 58% in chemistry, 88% in physics, and 79% in biology, closely rivals the best and newest closed reasoning model, Open I o1. For real-world expert-level tasks, *agentic Reasoning* was evaluated by domain experts, who noted that it effectively automated several hours of challenging, manual investigation. This highlights its potential to streamline labor-intensive processes and enhance productivity in knowledge-intensive domains.

Additionally, we tested the model’s scalability in test-time reasoning using the *agentic* framework as a verifier. The results showed significant improvements in test-time computational efficiency, demonstrating the framework’s ability to optimize reasoning processes. This finding suggests that the *agentic* framework has strong potential to serve as a reward model for reinforcement learning, further advancing reasoning model training.

These results position *agentic Reasoning* as a powerful and versatile framework, capable of tackling complex, domain-specific challenges with depth and precision. Its ability to perform in-depth research, navigate intricate logical structures, and synthesize information effectively highlights its potential for solving knowledge-intensive problems and driving advancements in deep analytical exploration.

## 2 Method

### 2.1 Preliminary

We consider an expert-level task that requires multi-step complex reasoning. In the process of model reasoning, it can retrieve external tool usage, and structured memory of its previous reasoning. Our objective is to generate, for each query  $q$ , both a logical reasoning chain  $r$  and a final answer  $a$ . To achieve this, the reasoning model dynamically interacts with external tools  $e$ , which are gener-

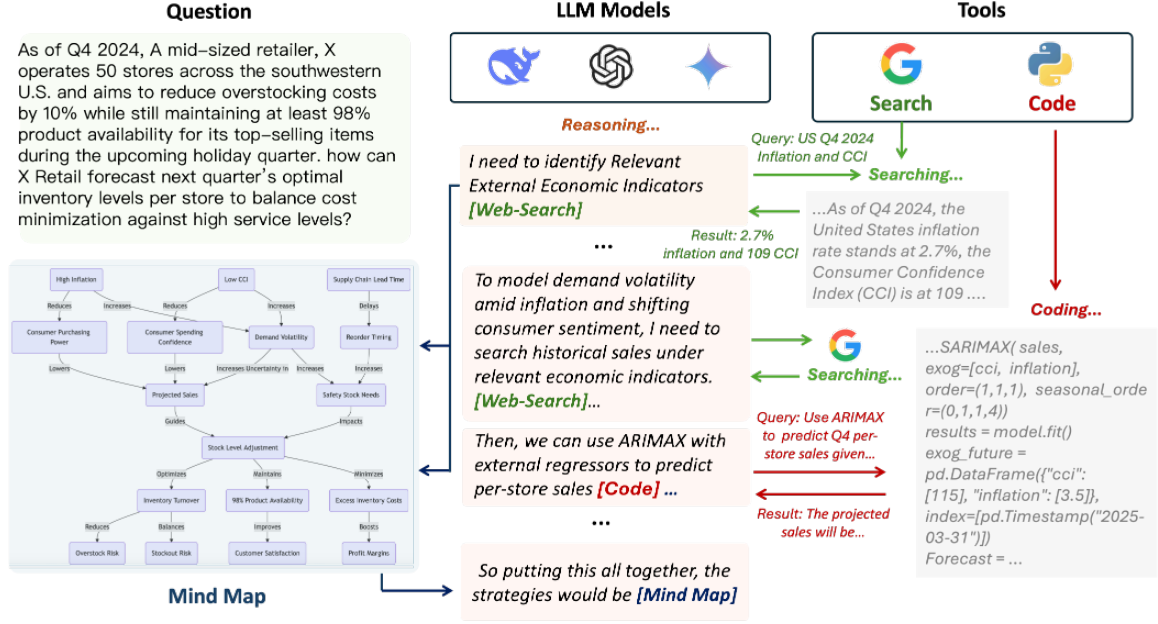


Figure 1: The overall workflow of agentic Reasoning.

ally web search and python coding, and retrieves structured knowledge from an organized memory  $k$  throughout the reasoning process.

Formally, we identify four primary inputs in the problem-solving pipeline: task instruction  $o$ , defining the overarching task objective, query  $q$ , a complex question requiring multi-step reasoning, external tool outputs  $e$ , dynamically retrieved content from tools such as web search or coding, reasoning memory  $k$ , containing structured knowledge graph.

The goal is to integrate  $o, q, e, k$  to generate a coherent reasoning chain  $r$  and a final answer  $a$ . This process can be expressed as the mapping:

$$(o, q, e, k) \mapsto (r, a)$$

We model the generation of  $r$  and  $a$  using the following joint probability formulation:

$$P(r, a | o, q, e, k) = \underbrace{\prod_{t=1}^{T_r} P(r_t | r_{1:t-1}, o, q, e_{1:t-1}, k_{1:t-1})}_{\text{Reasoning Process}} \underbrace{\prod_{t=1}^{T_a} P(a_t | a_{1:t-1}, r, o, q, e, k)}_{\text{Answer Generation}}$$

where  $T_r$  and  $T_a$  represent the lengths (in tokens) of the reasoning chain  $r$  and the final answer  $a$ , respectively. Here,  $r_t$  denotes the token at position  $t$  in the reasoning sequence, with  $r_{1:t}$  representing

all previous tokens. The terms  $e_t$  and  $k_t$  indicate all tool-generated outputs and knowledge-graph information retrieved up to step  $t$ . Similarly,  $a_t$  is the token at position  $t$  in the final answer, and  $a_{1:t}$  represents all previously generated answer tokens.

## 2.2 agentic Reasoning Pipeline

Our core idea is to enhance the model reasoning by deploying external LLM-based agents during reasoning. The framework enables the reasoning LLM model interacts with external information in an agentic way. During its reasoning process, it could call the external tools to help solve the problem and also with a structured memory, called Mind Map, to store its reasoning context. At its core, an agentic mechanism empowers the model to determine, in real-time, when additional information is required. Whenever the model identifies the external information is needed during its reasoning, it will proactively embed specialized tokens into its reasoning tokens. These tokens can be generally categorized to web-search token, coding token, and mind-map calling token. Together with token, the reasoning model would also generate a precise query as a message to interact with these external agents, based on the reasoning context developed so far.

Upon detecting such a token, the reasoning process temporarily halts to extract the query and its reasoning context. Those are then dispatched to external agents, such as search engines or Mind

Map, to generate pertinent content. The generation would consider both the message received and the reasoning context to make sure returning the most relevant results. These results are then reintegrated into the reasoning chain, allowing the model to continue its inference with an updated and enriched knowledge.

This iterative retrieval-and-reasoning cycle continues as needed, enabling the model to dynamically refine its conclusions until it reaches a fully reasoned final answer.

### 2.3 Mind Map gent

We construct a Mind Map to store and structure the real-time reasoning context of the reasoning model. This Mind Map is built by transforming raw reasoning chains into a structured knowledge graph. Specifically, we use a graph-construction LLM to extract entities from the reasoning chain and identify semantic relationships between related entities, following a process similar to that used in GraphR G (Edge et al., 2024).

The Mind Map serves two primary functions. First, it clusters reasoning context into distinct groups and summarizes each theme. This is achieved by applying community clustering (Edge et al., 2024) on the knowledge graph and using an LLM to generate concise summaries for each group. Second, the knowledge graph can be queried with specific questions, such as “Who was Jason’s maternal great-grandfather?” Using standard retrieval-augmented generation (R G) on the knowledge graph (Edge et al., 2024), we retrieve and return relevant information.

These functions integrate the Mind Map into various aspects of the gentic Reasoning process. It provides contextual reasoning support to external tools, enabling them to generate more context-aware responses (as discussed in later sections).

Additionally, when the reasoning model is uncertain about its claims or loses track in an extended reasoning process, it can query the Mind Map for relevant information, treating it as an external tool, and continue reasoning based on the retrieved answer.

### 2.4 Web-search gent

search agent is invoked to retrieve the most relevant documents from the web. Rather than incorporating the web pages in their raw form, they are temporarily held for further processing. This ensures that only the most pertinent information is

extracted and integrated into the main reasoning chain, maintaining coherence and relevance.

Once the relevant web pages are retrieved by the search agent, we use LLM to extract a concise, rephrased summary of the content most relevant to the ongoing reasoning context. This agent processes the web pages in the context of both the user query and the reasoning context, distilling key insights that are directly applicable to the problem at hand. The format and length of the summary adapt dynamically based on the reasoning task, for example, for factual queries like “What is the population of the US in 2024?” the result would be a simple numerical answer. For exploratory reasoning like finding a new perspective on a topic, the search agent would provide a summarized, detailed, nuanced viewpoint. For hypothesis validation like assessing supporting evidence for an assumption, the result would include the degree of support or contradiction found in the retrieved web-pages. This processed snippet is then integrated into the main reasoning process at the appropriate juncture, ensuring that external insights enhance rather than disrupt logical flow.

### 2.5 Coding gent

Instead of prompting the reasoning model to generate code directly, we find it more efficient to delegate coding tasks to a specialized coding LLM. The reasoning model sends the relevant context and query message to the coding LLM, which then writes the required code, executes it via a compiler, and returns the results. This approach ensures that the reasoning model remains focused on its core reasoning process without being disrupted by coding tasks, allowing for longer and more coherent reasoning chains. Specifically, we format the coding request as follows: “Write code to perform <code message from reasoning model> given the context <reasoning context from Mind Map> to answer the query <user query>.” The coding LLM is instructed to always return its output in natural language, ensuring seamless integration with the reasoning model.

### 2.6 Main Findings

**Less is More** Unlike general agentic frameworks that provide models with a large selection of external tools, we find that just two—web search and coding—are sufficient for most tasks, even those requiring expert-level proficiency. Adding more tools can degrade performance by increasing the



risk of inappropriate tool selection. Moreover, inaccuracies in external tool outputs can negatively impact the overall response quality. While additional tools are not significantly beneficial for language-based reasoning, they can be crucial for processing non-text modalities such as financial data, medical images, and genetic data. Developing specialized tools for different data modalities could further enhance LLM reasoning capabilities, and we will explore related results in future updates.

**Delegating Tasks to LLM-Based Agents** Distributing computational workloads across multiple LLM-based agents improves efficiency. Instead of having the main reasoning model handle all tool-related tasks (e.g., writing code or constructing a knowledge graph), or calling non-LLM tools like pure search engine or code compiler, we delegate these tasks to specialized LLM-Based agents, like a coding LLM generates code based on the query and context from the main reasoning model, or a knowledge-graph LLM constructs structured representations (e.g., a Mind Map) from the reasoning chain. This approach offers two key advantages: 1. **Minimizing Disruptions.** The main reasoning model can maintain longer, more coherent reasoning without being distracted by auxiliary tasks or exceeding token limits. 2. **Leveraging Specialization.** Different LLMs excel at different tasks—for instance, DeepSeek-R1 specializes in reasoning, while Claude-Sonnet excels at coding. By assigning tasks to models best suited for them, we achieve higher overall performance.

**Agentic Test-time Scaling?** For a single question, we find reasoning chains that utilize more tool calls tend to yield better results. While across different questions, those requiring excessive tool usage often indicate inherent ambiguity or inaccuracy in the initial reasoning. This insight can be leveraged as a test-time reasoning verifier. By selecting the reasoning chain with the highest tool usage, we can implement best-of-N selection or beam search, which are techniques commonly used in mathematical and coding reasoning tasks as they can easily build a verifier, to open-domain, knowledge-intensive Q&A, improving accuracy and robustness.

## Question

A 68-year-old male with a 12-year history of moderate-to-severe COPD ( $FEV_1 = 45\%$  of predicted,  $PaO_2 = 58$  mmHg,  $PaCO_2 = 48$  mmHg) and newly diagnosed heart failure (left ventricular ejection fraction,  $LVEF = 35\%$ ) presents with worsening dyspnea and reduced exercise tolerance. how can lung function be improved without exacerbating heart symptoms?

## Reasoning...

**[Code]** Target an Alveolar  $PAO_2$  of  $\sim 140$  mmHg, anticipating an A-a gradient that might still yield an arterial  $PaO_2 \geq 65 \geq 65$  mmHg using Alveolar Gas Equation:  $PAO_2 = FiO_2 \times (P_{atm} - P_{H_2O}) - \frac{PaCO_2}{R}$

Thus, **28%  $FiO_2$**  might be a reasonable starting point. In practice, clinicians often start with **24–28%** supplemental  $O_2$  via a Venturi mask (or nasal cannula approximations) and then titrate to keep  **$SpO_2$  around 88–92%** in COPD to avoid excessive respiratory drive suppression or severe hypercapnia.

...

**[Web-search]** Optimal PEEP levels in COPD with heart failure  
Low-to-moderate **PEEP (5–10 cm  $H_2O$ )** is recommended to improve oxygenation without significantly reducing venous return in HF.

...

Administer  $\sim 28\%$   $FiO_2$ , consider mild PEEP/CPAP of 4–5 cm  $H_2O$ , increase alveolar ventilation modestly, and titrate diuretics to manage fluid overload without excessive preload reduction.

Figure 2: Case study on a complex medical decision-making problem.

## 3 Experiments

### 3.1 Solving Hard Problems

We evaluate our agentic Reasoning model on the GPQ dataset, a PhD-level multiple-choice science Q&A benchmark. The dataset consists of expert-authored questions spanning physics, chemistry, and biology. Our primary experiments focus on the high-quality Diamond Set, which contains 198 questions, while Table 2 presents results on the broader Extended Set of 546 questions, allowing for a direct comparison with human experts.

As shown in Table 1, our findings show that large reasoning models such as DeepSeek-R1-Lite and

Table 1: Performance comparison on GPQ dataset across Physics, Chemistry, and Biology.

Method	Phy.	Chem.	Bio.
<i>Direct Reasoning</i>			
Qwen2.5-32B	57.0	33.3	52.6
Qwen2.5-Coder-32B	37.2	25.8	57.9
QwQ-32B	75.6	39.8	68.4
Qwen2.5-72B	57.0	37.6	68.4
Llama3.3-70B	54.7	31.2	52.6
GPT-4o <sup>†</sup>	59.5	40.2	61.6
o1-preview <sup>†</sup>	89.4	59.9	65.9
<i>Retrieve/Search in Reasoning</i>			
R G-Qwen2.5-32B	57.0	37.6	52.6
R G-QwQ-32B	76.7	38.7	73.7
R gent-Qwen2.5-32B	58.1	33.3	63.2
R gent-QwQ-32B	76.7	46.2	68.4
Search-o1	77.9	47.3	78.9
<i>genic Reasoning</i>			
Ours	<b>88.1</b>	<b>58.3</b>	<b>79.6</b>

QwQ-32B-Preview significantly outperform traditional instruction-tuned LLMs. This demonstrates the effectiveness of chain-of-thought reasoning in solving complex, expert-level problems. Additionally, models like R gent-QwQ-32B and Search-O1, which autonomously retrieve relevant information at reasoning, outperform non-reasoning models that simply utilize search tools. This confirms that calling tools is uniquely beneficial for enhancing reasoning accuracy.

genic Reasoning, which integrates external agents during reasoning, further improves performance over search-enhanced models. Our model achieves superior results on the GPQ dataset, demonstrating the power of tool-assisted reasoning in tackling expert-level challenges.

To illustrate the effectiveness of genic Reasoning, we also present a case study on a complex medical decision-making problem, as shown in Figure 4. The model autonomously executes code to compute the optimal  $FiO_2$  (Fraction of Inspired Oxygen) for a patient, performs a web search to retrieve the most accurate PEEP (Positive End-Expiratory Pressure) value, and synthesizes both results to determine the best treatment plan. This example highlights how integrating coding and web search enhances the model’s ability to solve real-world medical challenges.

We further compare our model with human experts in physics, chemistry, and biology using the

GPQ Extended Set. As shown in Table 2, our model surpasses human performance across all disciplines, achieving superior accuracy in all three subsets, and also outperforming human experts. These results highlight the model’s ability to handle specialized scientific reasoning tasks at an expert level.

Table 2: Performance comparison with human experts on the GPQ extended set.

Method	Phy.	Chem.	Bio.
<i>Human Experts</i>			
Physicists	57.9	31.6	42.0
Chemists	34.5	72.6	45.6
Biologists	30.4	28.8	68.9
<i>Reasoning Models</i>			
QwQ-32B	61.7	36.9	61.0
R G-QwQ-32B	64.3	38.3	66.7
Search-o1	68.7	40.7	69.5
genic Reasoning	75.2	53.1	72.8

## 3.2 Deep Research

We conduct an evaluation of genic Reasoning for deep research in open-ended Q&A tasks. A group of PhD-level experts in finance, medicine, and law were asked to formulate 15 to 30 professional research questions closely related to their respective fields. These questions were designed to require at least 20 minutes of in-depth research to answer comprehensively.

We assess the accuracy and reliability of reports generated by our genic Reasoning model, measuring the pass rate—the percentage of responses deemed satisfactory by domain experts. We compare this pass rate against Gemini Deep Research Service (experiments with OpenAI’s Deep Research are ongoing). As shown in Figure 3, our findings show that genic Reasoning outperforms Gemini Deep Research across all three domains, demonstrating the effectiveness of structured reasoning and tool-augmented frameworks in conducting deep research.

## 3.3 Analysis

### 3.3.1 Test-time Scaling

In our deep research study, we find that increased tool usage improves performance on the same question. As shown in Figure 3, a higher number of tool calls by the reasoning model correlates with an increased pass rate in deep research tasks. How-

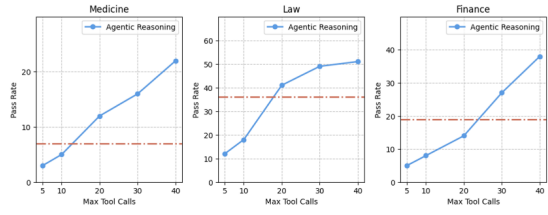


Figure 3: More calling for agentic tools, the better the model does. Red line denotes Gemini Deep Research

ever, when comparing different questions, those requiring excessive tool usage tend to indicate an inherently more challenging or ambiguous question, leading to lower accuracy. The questions with a higher number of tool calls within the same field ultimately achieve a lower pass rate.

Such observations provide a practical approach for test-time scaling. During inference-time search (running the same question multiple times), we can use the frequency of tool calls as a heuristic to select better responses. simple implementation, such as best-of-N selection, can effectively filter out weaker outputs. This method even outperforms LLM-as-a-judge evaluation, which is more computationally expensive, time-consuming, and prone to instability.

These findings suggest a promising direction for reinforcement learning for reasoning model on knowledge-intensive fields. By leveraging agentic tool usage as an implicit reward signal, we can further optimize reasoning models for more effective tool utilization, ultimately enhancing their problem-solving capabilities.

### 3.3.2 The Role of Mind Map

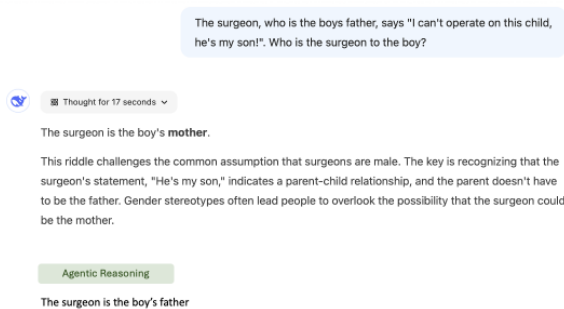


Figure 4: tricky question that misleads most LLMs is correctly answered by us.

We find that Mind Maps are particularly effective in clarifying complex logical relationships, enabling the model to solve problems that often mislead traditional LLMs. We highlight two key cases where Mind Mapping maximizes its capabilities:

First, Mind Maps help correctly answer tricky logic-based questions that frequently fool LLMs. well-known example is a modified riddle: "The surgeon, who is the boy's father, says 'I can't operate on this child, he's my son!' Who is the surgeon to the boy?" DeepSeek-R1 took 17 seconds to process this question but still produced the wrong answer, a failure also observed in models from the GPT and Gemini series models. These models often fall for a political-correct corpus contaminated response, failing to recognize the obvious logical structure. However, in our agentic Reasoning framework, the use of a Mind Map allows the model to explicitly analyze the logical relationships between the entities [surgeon], [boy], and [father], leading to the correct answer.

Second, Mind Maps enhance deductive reasoning in strategic games. We test our approach in Werewolf, a classic social deduction game where players take on hidden roles as either villagers or werewolves. Villagers attempt to identify the werewolves, while werewolves deceive the group and eliminate players without being caught. The game alternates between "night", where werewolves secretly attack, and "day", where players debate and vote on eliminations. To evaluate our agentic Reasoning model, we invited seven experienced Werewolf players (5+ years of experience) to play against it. The model achieved an impressive 72% win rate, significantly exceeding both the expected statistical win rate and the performance of human players in our experiment.

We analyzed the Mind Maps generated by the agentic Reasoning model over multiple rounds of play, as shown in Figure 5. These visual structures helped the model track the relationships between different players based on their spoken arguments, allowing it to more accurately identify deception strategies, anticipate voting behaviors, and optimize its own disguise tactics. This result demonstrates that Mind Mapping is not just a tool for logic puzzles but also a powerful strategy enhancer in dynamic reasoning environments.

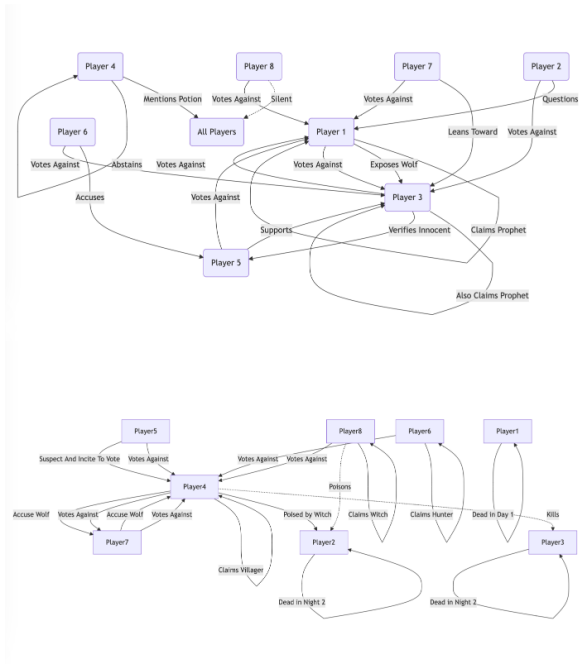


Figure 5: Mind Map in playing werewolf game. The first round and the second round.

## 4 Conclusion

We introduced `gentic Reasoning`, a framework that enhances LLM reasoning by integrating external agents for structured memory (Mind Map), web search, and computational analysis. This approach improves logical coherence, factual accuracy, and deep research capabilities. Our evaluations show that `gentic Reasoning` outperforms existing models on expert-level Q&A and real-world research tasks, demonstrating its ability to synthesize knowledge effectively. The structured use of external tools enables more interpretable and verifiable reasoning, paving the way for AI systems capable of expert-level problem-solving. Future work will explore extending this framework to multimodal data and real-time adaptability, further advancing LLaMA’s ability to tackle complex, real-world challenges.

## References

- Figure 5: Mind Map in playing werewolf game. The first round and the second round.

Zhen Huang, Haoyang Zou, Xuefeng Li, Yixiu Liu, Yuxiang Zheng, Ethan Chern, Shijie Xia, Yiwei Qin, Weizhe Yuan, and Pengfei Liu. 2024. O1 replication journey—part 2: Surpassing o1-preview through simple distillation, big progress or bitter lesson? *arXiv preprint arXiv:2411.16489*.

Shayekh Bin Islam, Md Sib Rahman, KSM Hossain, Enamul Hoque, Shafiq Joty, and Md Rizwan Parvez. 2024. Open-rag: Enhanced retrieval-augmented reasoning with open-source large language models. *arXiv preprint arXiv:2410.01782*.

aron Jaech, dam Kalai, dam Lerer, dam Richardson, hmed El-Kishky, iden Low, lec Helyar, leksander Madry, lex Beutel, lex Carney, et al. 2024. Openai ol system card. *arXiv preprint arXiv:2412.16720*.

M Abdul Khaliq, P Chang, M Ma, Bernhard Pflugfelder, and F Miletic. 2024. Ragar, your falsehood radar: Rag-augmented reasoning for political fact-checking using multimodal large language models. *arXiv preprint arXiv:2404.12065*.

itor Lewkowycz, nders ndreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, mbrose Slone, Cem nil, Imanol Schlag, Theo Gutman-Solo, et al. 2022. Solving quantitative reasoning problems with language models. *dvances in Neural Information Processing Systems*, 35:3843–3857.

Xiaoxi Li, Guanting Dong, Jiajie Jin, Yuyao Zhang, Yujia Zhou, Yutao Zhu, Peitian Zhang, and Zhicheng Dou. 2025. Search-o1: genetic search-enhanced large reasoning models. *arXiv preprint arXiv:2501.05366*.

Open I. Learning to reason with LLMs.

Yiwei Qin, Xuefeng Li, Haoyang Zou, Yixiu Liu, Shijie Xia, Zhen Huang, Yixin Ye, Weizhe Yuan, Hector Liu, Yuanzhi Li, et al. 2024. O1 replication journey: strategic progress report—part 1. *arXiv preprint arXiv:2410.18982*.

Yijia Shao, Yucheng Jiang, Theodore Kanell, Peter Xu, Omar Khattab, and Monica S Lam. 2024. *Assisting in writing wikipedia-like articles from scratch with large language models. arXiv preprint arXiv:2402.14207.*

DeepSeek Team. 2024. Deepseek-r1-lite-preview is now live: unleashing supercharged reasoning power.

Qwen Team. Qwq: Reflect deeply on the boundaries of the unknown, november 2024. URL <https://qwenlm.github.io/blog/qwq-32b-preview>.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *dvances in neural information processing systems*, 35:24824–24837.

Di Zhang, Jianbo Wu, Jingdi Lei, Tong Che, Jiatong Li, Tong Xie, Xiaoshui Huang, Shufei Zhang, Marco Pavone, Yuqiang Li, et al. 2024. Llama-berry: Pairwise optimization for o1-like olympiad-level mathematical reasoning. *arXiv preprint arXiv:2410.02884*.