Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

SearchExpert: A GenAI-driven framework for reasoning-intensive multimedia information fusion through fine-tuning and reinforcement learning

Jinzheng Li [a], Yiqing Shen [b], Wei Zhou [c], Hui Chen [d],*

[a] School of Computer Science and Technology, Fudan University, Shanghai, 200433, China
[b] Johns Hopkins University, Baltimore, 21218, USA
[c] Cardiff University, Cardiff, CF10 3AT, UK
[d] Macquarie University, Sydney, 2109, Australia

ARTICLE INFO

Keywords:
Large language models
Complex reasoning
Multimodal search
Reinforcement learning
DAG planning
Vision-language models

ABSTRACT

The rapid advancement of Generative Artificial Intelligence (GenAI) has opened new frontiers in multimodal information fusion, yet current large language model (LLM)-driven search agents remain limited in their ability to handle reasoning-intensive queries and integrate multimedia data effectively. In this paper, we propose SearchExpert, a GenAI-enhanced framework that augments LLMs with powerful multimedia search and reasoning capabilities via a novel two-stage training paradigm. First, we introduce an efficient natural language representation for directed acyclic graph (DAG)-based search plans to reduce token overhead and support structured reasoning. We then propose Supervised Fine-Tuning for Searching (SFTS), enabled by an automated data construction pipeline that adapts LLMs to generate token-efficient, structured search plans from complex queries. Second, to further enhance reasoning ability, we introduce Reinforcement Learning from Search Feedback (RLSF), which uses reward signals based on semantic alignment and intrinsic quality assessments of retrieved results to optimize LLM behavior. To address the limitations of unimodal input and output, we integrate a multimedia understanding and generation module based on vision-language models and image synthesis tools (e.g., BLIP-2, and DALLE-3), enabling the GenAI-based fusion of text and visual data. We also establish SearchExpertBench-25, a benchmark comprising 200 multimedia-rich, reasoning-intensive queries spanning financial and global news domains, accompanied by a rigorous human evaluation framework. Experimental results demonstrate that SearchExpert surpasses state-of-the-art baselines such as FinSearch and Perplexity Pro, achieving up to 71.5% accuracy on complex benchmark tasks while reducing token consumption by over 40%. Human evaluations further highlight improvements in completeness, analytical integrity, and multimodal fluency. This work presents a scalable and generalizable GenAI framework for information fusion, with implications for real-time decision-making in complex, multi-source environments. The code is available at https://anonymous.4open.science/r/SearchExpert-2343/.

1. Introduction

Large language models (LLMs) have transformed search interactions by enabling users to express complex natural language queries beyond simple keywords [1–4]. Specifically, existing agent-based search methods such as MindSearch [5], FinSearch [6], and MMSearch [7] leverage LLMs to decompose user queries into a directed acyclic graph (DAG) represented search plans, where nodes represent specific search keywords (and the corresponding search source), while edges denote logical dependencies between them. However, MindSearch [5] suffers from pre-determined search plans and temporal insensitivity. Subsequently, FinSearch [6] improved MindSearch with adaptive planning graphs and time-sensitive response generation for financial-related search queries, but remains limited in multimodal information processing [8–10]. Although MMSearch [7] incorporates visual input by converting them into text, it struggles to generate multimedia output. Most importantly, all these existing agent-based search methods rely on prompt engineering to decompose queries [11,12], which restricts their effectiveness in complex search scenarios involving reasoning. Moreover, they also suffer from high excessive token consumption due to Python-based DAG representations, which limits practical efficiency [13].

* Corresponding author.
E-mail address: huichen.ml@gmail.com (H. Chen).

https://doi.org/10.1016/j.inffus.2025.103665
Received 18 June 2025; Received in revised form 23 July 2025; Accepted 26 August 2025
Available online 1 September 2025

Table 1
Comparison of LLM-driven search methods across major capabilities. "**Efficient Representation**" indicates whether the method employs token-efficient natural language representations for search plans rather than code-based DAG implementations. "**SFT**" shows if the method incorporates supervised fine-tuning specifically optimized for search planning. "**RL**" indicates whether the method leverages reinforcement learning to enhance searching or reasoning capabilities. "**Multimedia Integration**" refers to the ability to process visual inputs and generate visual elements in responses. "**Complex Query Handling**" demonstrates the capability to effectively process reasoning-intensive queries with multiple variables and deliberate obfuscation.

Methods	Efficient representation	SFT	RL	Multimedia integration	Complex query handling
MindSearch [5]	×	×	×	×	×
FinSearch [6]	×	×	×	✓	×
MMSearch [7]	×	×	×	✓	×
SearchExpert (Ours)	✓	✓	✓	✓	✓

To address these limitations, we introduce SearchExpert, a two-stage training framework that enhances LLMs' multimedia search capabilities through three following innovations. First, we introduce a novel natural language representation for DAG-based search plans, which replaces traditional Python-style code representations. This design not only significantly reduces token usage—over 40% in our benchmarks—but also improves interpretability and model alignment. Unlike prior work that encodes search plans as verbose code-like graphs, our approach offers a token-efficient yet structurally faithful textual format tailored for LLM generation. Second, to improve reasoning capabilities for complex search queries, we propose reinforcement learning from search feedback (RLSF). Based on RLAIF [14], RLSF uses LLM to evaluate the quality of the search results and transforms these evaluations into reinforcement signals to update the LLM through Proximal Policy Optimization (PPO) [15]. Our approach uniquely assesses both the semantic similarity to reference and the intrinsic quality of search results. Third, we integrate a multimedia understanding and generation agent that enables SearchExpert to process visual inputs and generate visual outputs during inference, allowing for the handling of multimedia queries [16]. Table 1 shows the high-level difference of our method from previous work. Compared to FinSearch's adaptive but fixed-format query plans, our method generates open-ended natural language DAGs with flexible dependency structures. Unlike MindSearch, which relies on hard-coded DAG modules, we treat DAG generation as a language modeling task and optimize it via reinforcement learning. Moreover, we extend the planning framework to handle multimodal queries through vision-language integration, a capability absent in both prior systems.

In this work, we introduce **SearchExpert**, a GenAI-driven framework for reasoning-intensive multimedia information fusion, featuring a two-stage learning paradigm:

- A natural-language DAG format encodes search and fusion steps, cutting token usage by over 40% without sacrificing structural clarity.
- An automated data pipeline generates high-quality query–plan pairs for supervised fine-tuning, followed by reinforcement learning with a dual-component reward – semantic alignment plus intrinsic quality – to sharpen reasoning and retrieval via PPO.
- Vision–language (BLIP-2) and image synthesis (DALLE-3) modules enable seamless text–visual integration, and SearchExpertBench-25 (200 finance/news queries) with human evaluation assesses completeness, timeliness, analytical integrity, and fluency.

The rest of this paper is organized as follows. Section 2 formalizes the problem and details the SearchExpert methodology, including efficient plan representation, SFTS, RLSF, and multimodal integration. Section 3 describes the experimental setup, benchmark construction, implementation specifics, and empirical results. Finally, Section 4 concludes and outlines avenues for future research.

2. Methods

2.1. Problem formulation

We formalize the reasoning-intensive multimedia search with LLMs as a process of generating search plans for complex queries. Given a user query q that contains both textual and visual elements, the LLM first generates a structured search plan S represented as a directed acyclic graph (DAG) $G = (V, E)$, formalized as $G = \texttt{LLM}(q)$. The major focus of this work is to enhance the LLM's capability to generate effective search plans through our proposed SFTS and RLSF training, particularly for scenarios requiring complex reasoning. In this formulation, $V = \{v_1, v_2, \dots, v_n\}$ represents the set of search nodes, with each node containing specific keywords for the search along with the designated search source (*e.g.*, general search, news search, financial search). The edge set $E \subseteq V \times V$ represents the logical dependencies between the search steps, where an edge (v_i, v_j) indicates that the search results of node v_i directly inform the query formulation for node v_j. To ensure a coherent and feasible search strategy, the graph must maintain acyclicity, preventing any sequence $v_{i_1} \to v_{i_2} \to \cdots \to v_{i_k} \to v_{i_1}$ from forming a cycle in G. A major innovation in our approach is to represent G in an efficient natural language format instead of the Python-based implementations used in previous work [5], which reduces token consumption and processing overhead. To make the transformation clearer, consider a simple example: A traditional Python-style DAG may define two dependent steps: first searching "Japan earthquake" in news sources, then using its results to search "supply chain impact" in financial sources. In our natural language representation, this becomes: *"A: Japan earthquake (news) B: Supply chain impact (finance) A → B"*. This format preserves the structure and logic of the original DAG, while significantly reducing token usage. Specifically, each node $v_i \in V$ is expressed as "v_i: node feature", where the node feature comprises search keywords and the corresponding search source. The edges E are represented as directional relationships between nodes, expressed as "$v_i \to v_j$". The complete representation can be formalized as follows:

$$\text{Rep}(G) = \{(v_i : k_i) \mid v_i \in V, k_i \text{ is the keyword and search source}\} \cup \{(v_i \to v_j) \mid (v_i, v_j) \in E\}. \tag{1}$$

For example, a search plan investigating economic impacts might be represented as "*A: Japan earthquake impact (news search), B: global supply chain disruption (financial search), C: automotive industry response (industry search)... A → B, B → C*". This indicates a sequential search process where each node's results inform subsequent queries, enabling multi-step reasoning through structured information gathering. After generation, the search plan S is executed to retrieve relevant information, which is then synthesized into a final response R by another frozen LLM. This two-stage formulation separates search planning from

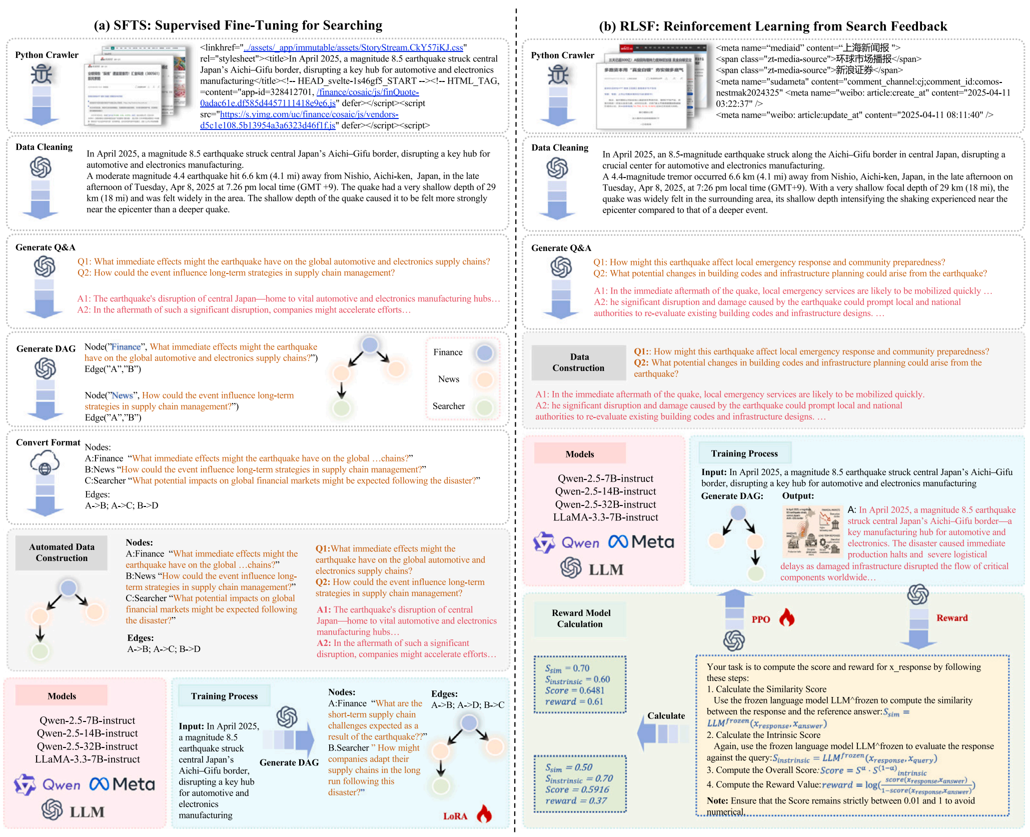


Fig. 1. The two-stage training framework of SearchExpert. (a) Supervised fine-tuning for searching (SFTS): Our automated data construction pipeline begins with Python crawlers collecting recent online content, followed by data cleaning and Q&A generation. We then generate efficient natural language DAG representations of search plans, converting from traditional Python-based formats to our token-efficient representation. The constructed training data is used to fine-tune LLMs to generate structured search plans directly from queries. (b) Reinforcement searching from search feedback (RLSF): Building upon SFTS, our RLSF approach uses similar initial data collection but focuses on constructing more complex, reasoning-intensive search scenarios. The trained models generate search plans that are executed to retrieve information, with search quality evaluated through our dual-component reward mechanism combining semantic similarity and intrinsic quality assessment. The resulting scores guide LLM optimization through PPO.

response synthesis, allowing us to focus our SFTS and RLSF training specifically on optimizing the search planning capabilities of the LLM, which directly impacts the quality and relevance of the final response. We define a reasoning-intensive query as one that requires not only multi-hop information retrieval but also explicit logical inference, causal reasoning, or temporal alignment across heterogeneous information sources. Unlike standard complex or multi-hop queries, which may involve multiple steps of retrieval, reasoning-intensive queries demand the integration, synthesis, and interpretation of intermediate results to reach a non-obvious conclusion. For instance, assessing the long-term economic implications of a regional disaster involves not just retrieving facts, but inferring macroeconomic trends from cross-domain evidence.

2.2. Supervised fine-tuning for searching

The goal of our SFTS is to train LLMs to generate the efficient natural language DAG representation $\mathrm{Rep}(G)$ defined in Eq. (1) when presented with a complex search query q. Unlike standard supervised fine-tuning approaches that rely on manually constructed datasets and generic optimization objectives, SFTS leverages an automated data construction pipeline specifically designed for search planning, as shown in Fig. 1. While the natural language DAG representation offers substantial token savings and improves readability, it may introduce minor ambiguity in edge cases with deeply nested dependencies or similarly named nodes. However, in our experiments, such ambiguity was rare and did not significantly impact plan execution or retrieval quality. In future work, hybrid representations combining natural language with lightweight structural markers may further enhance expressiveness.

2.2.1. Automated data construction

Our automated data construction pipeline reduces the manual effort typically required for SFT while ensuring high-quality training data for search-specific scenarios. It begins by crawling recent online texts x_{online} from various time-sensitive sources, including news outlets, financial publications, and government websites. Focusing on current information ensures that the corresponding content does not appear in the base LLM's training corpus, thereby necessitating search capabilities rather than knowledge retrieval. For each crawled text, we employ a frozen LLM to generate query-answer pairs, namely $(x_{\text{query}}, x_{\text{answer}}) = \mathrm{LLM}^{\text{frozen}}(x_{\text{online}})$. To verify that these pairs require actual search capabilities rather than relying on information already present in the LLM's knowledge, we filter it by presenting each query x_{query} to a separate LLM trying to answer it without access to the corresponding online context x_{online} and analyze its response, *i.e.*

$$\mathrm{isNovel}(x_{\text{query}}) = \begin{cases} \text{True}, & \text{if } \mathrm{Sim}(\mathrm{LLM}^{\text{frozen}}(x_{\text{query}}), x_{\text{answer}}) < \tau \\ \text{False}, & \text{otherwise} \end{cases} \tag{2}$$

where $\mathrm{Sim}(\cdot,\cdot)$ measures the semantic similarity determined by LLM, and τ represents a threshold that determines sufficient novelty. This filtering step again distinguishes our method from standard SFT approaches by ensuring that the training data specifically targets search-dependent queries rather than general knowledge queries. Next, we feed each qualified query x_{query} into FinSearch to generate both a search plan encoded in Python format G_{Python} and the corresponding final response x_{response}. We take an additional step of validating response quality before converting the to G_{Python} our efficient representation:

$$\text{isAligned}(x_{\text{response}}, x_{\text{answer}}) = \begin{cases} \text{True}, & \text{if } \text{Sim}(x_{\text{response}}, x_{\text{answer}}) > \delta \\ \text{False}, & \text{otherwise} \end{cases} \tag{3}$$

where δ is the alignment threshold. For aligned responses, we transform the Python-based search plan representation into our natural language representation described in Eq. (1) through $\text{Rep}(G) = \text{LLM}^{\text{frozen}}(G_{\text{Python}})$ to reduce the number of tokens that represent G while preserving the logical structure of the search plan. The final output comprises pairs of queries and their corresponding representations of the natural language search plan $x_{\text{train}} = \{$ "input": x_{query}, "output": $\text{Rep}(G)\}$.

2.2.2. Training process

Using the automated generated dataset in the format $x_{\text{train}} = \{$ "input": x_{query}, "output": $\text{Rep}(G)\}$, we train the LLM to generate natural language DAG representations directly from queries, expressed as:

$$\text{Rep}(G)_{\text{pred}} = \text{LLM}_\theta(x_{\text{query}}), \tag{4}$$

where LLM_θ represents our target LLM with parameters θ, and $\text{Rep}(G)_{\text{pred}}$ is the predicted natural language representation of the search plan. The optimization objective employs a standard cross-entropy loss function that aligns the outputs with the reference natural language DAG representations.

2.3. Reinforcement learning from search feedback

While SFTS provides a foundation for generating efficient search plans with LLM, complex reasoning-intensive queries often require capabilities beyond what supervised learning alone can achieve [17–19]. Our RLSF addresses this limitation through a novel reward feedback mechanism that directly optimizes for search result quality rather than merely imitating reference search plans [20]. While our RLSF approach focuses on optimizing the generation of search plans (i.e., DAG structures), reinforcement learning could also be applied post hoc—for example, to re-rank or filter retrieved content based on downstream quality metrics. However, we find that tightly coupling RL with the planning stage provides greater upstream control over information gathering. By guiding the model to produce higher-quality search structures upfront, we reduce reliance on post-retrieval heuristics and ensure that relevant content is discovered rather than merely re-ranked. This also aligns better with our goal of improving reasoning performance through better planning rather than retroactive correction.

2.3.1. Automated dataset construction

We construct the RLSF dataset using methods similar to SFTS, but with heightened emphasis on complex reasoning scenarios that require multi-step search processes. Unlike SFTS which relies on single-source crawled content, RLSF dataset construction employs diverse and dynamic sources to generate more challenging time-sensitive queries, which ensures exposure to a broader range of reasoning patterns and information dependencies. For RLSF training, we only need the query-answer pairs $(x_{\text{query}}, x_{\text{answer}})$ rather than predetermined search plans, as optimization is driven by the quality of the search outcome rather than the adherence to the structure of the search plan.

2.3.2. Training process

RLSF training begins with the LLM that generates a representation of a natural language search plan based on the input query: $\text{Rep}(G)_{\text{pred}} = \text{LLM}_\theta(x_{\text{query}})$. This search plan is executed to retrieve information and generate the corresponding response:

$$x_{\text{response}} = \texttt{execute}(\text{Rep}(G)_{\text{pred}}). \tag{5}$$

The major innovation in our RLSF is the dual component reward calculation that captures both alignment with reference answers and intrinsic quality metrics. For semantic similarity assessment, we prompt an LLM to compare the generated response with the reference answer:

$$S_{\text{sim}} = \texttt{LLM}^{\text{frozen}}(x_{\text{response}}, x_{\text{answer}}), \tag{6}$$

where LLM is prompt to analyze the factual consistency, the coverage of information, and the contextual relevance between the two texts, producing a normalized similarity score between 0 and 1. This component ensures that the search plan leads to responses that capture the essential information needed to address the query. Complementing this, our intrinsic quality assessment employs another LLM assessment that evaluates the completeness, precision, coherence, and clarity of the response independently of any reference.

$$S_{\text{intrinsic}} = \texttt{LLM}^{\text{frozen}}(x_{\text{response}}, x_{\text{query}}). \tag{7}$$

These two factors are combined through a weighted geometric mean that balances fidelity to reference answers with intrinsic search quality:

$$\text{score} = S_{\text{sim}}^{\alpha} \cdot S_{\text{intrinsic}}^{(1-\alpha)}, \tag{8}$$

where $\alpha \in [0, 1]$ controls the relative importance of similarity versus intrinsic quality. We use a geometric mean rather than an additive model to ensure that both reward components – semantic similarity and intrinsic quality – have multiplicative influence on the final score. This formulation penalizes cases where one component is low, encouraging balanced improvements across both dimensions. The resulting score is transformed into a reinforcement learning signal using a log-odds transformation:

$$\text{reward} = \log\left(\frac{\text{score}(x_{\text{response}}, x_{\text{answer}})}{1 - \text{score}(x_{\text{response}}, x_{\text{answer}})}\right). \tag{9}$$

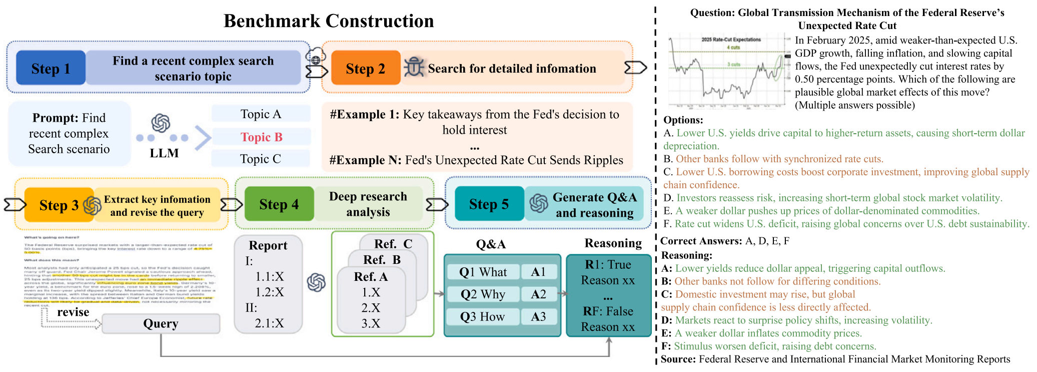
This reward signal guides the optimization of LLM through PPO [15]. We apply a log-odds transformation to both the similarity and quality scores to map probabilities into the real number space and magnify differences near decision boundaries. This transformation improves gradient signal strength when scores are near 0.5, which stabilizes RL training and enhances sensitivity to model improvements. Additionally, log-odds provide more interpretable reward scaling compared to raw probabilities.

2.4. Multimedia understanding and generation

While SFTS and RLSF enhance the LLM's ability to generate effective search plans and reason across complex information sources, handling multimedia content requires additional operation for both input understanding and output generation. Our multimedia agent integrates visual processing and generation capabilities into the SearchExpert. To process multimodal queries containing visual elements, we implement a visual understanding pipeline based on BLIP-2 [21], a vision-language model that converts visual information into textual representations. Given a query q that contains both textual components q_{text} and visual components q_{visual}, our agent converts visual elements into textual descriptions $q_{\text{visual\_desc}} = \texttt{BLIP}-2(q_{\text{visual}})$. The resulting textual description is then merged with the original textual query to form an enriched query *i.e.*, $q_{\text{enriched}} = \texttt{merge}(q_{\text{text}}, q_{\text{visual\_desc}})$. This enriched query is processed by our fine-tuned LLM to generate the natural language search plan $\text{Rep}(G)$. To improve response quality and information comprehensiveness, we incorporate a graph generation that produces visual elements based on textual descriptions via DALLE-3 [22], represented as $v_{\text{output}} = \texttt{DALLE} - 3(g_{\text{desc}})$, where g_{desc} is a textual description of the desired graphical output derived from search results, and v_{output} is the generated visual element.

2.5. Benchmark dataset construction

To evaluate SearchExpert's reasoning-intensive multimedia search capabilities, we developed SearchExpertBench-25, a benchmark comprising 200 complex multiple-choice questions spanning financial and


Benchmark Construction
Step 1
Find a recent complex search scenario topic
Step 2
Search for detailed infomation
Prompt: Find recent complex Search scenario
LLM
Topic A
Topic B
Topic C
#Example 1: Key takeaways from the Fed's decision to hold interest
...
#Example N: Fed's Unexpected Rate Cut Sends Ripples
Step 3
Extract key infomation and revise the query
Step 4
Deep research analysis
Step 5
Generate Q&A and reasoning
revise
Query
Report
I:
1.1:X
1.2:X
II:
2.1:X
Ref. C
Ref. B
Ref. A
1.X
2.X
3.X
Q&A
Q1 What
A1
Q2 Why
A2
Q3 How
A3
Reasoning
R1: True
Reason xx
...
RF: False
Reason xx
2025 Rate-Cut Expectations
4 cuts
3 cuts

Question: Global Transmission Mechanism of the Federal Reserve's Unexpected Rate Cut

In February 2025, amid weaker-than-expected U.S. GDP growth, falling inflation, and slowing capital flows, the Fed unexpectedly cut interest rates by 0.50 percentage points. Which of the following are plausible global market effects of this move? (Multiple answers possible)

Options:
A. Lower U.S. yields drive capital to higher-return assets, causing short-term dollar depreciation.
B. Other banks follow with synchronized rate cuts.
C. Lower U.S. borrowing costs boost corporate investment, improving global supply chain confidence.
D. Investors reassess risk, increasing short-term global stock market volatility.
E. A weaker dollar pushes up prices of dollar-denominated commodities.
F. Rate cut widens U.S. deficit, raising global concerns over U.S. debt sustainability.

Correct Answers: A, D, E, F
Reasoning:
A: Lower yields reduce dollar appeal, triggering capital outflows.
B: Other banks not follow for differing conditions.
C: Domestic investment may rise, but global supply chain confidence is less directly affected.
D: Markets react to surprise policy shifts, increasing volatility.
E: A weaker dollar inflates commodity prices.
F: Stimulus worsen deficit, raising debt concerns.
Source: Federal Reserve and International Financial Market Monitoring Reports

Fig. 2. Benchmark construction method and example case from SearchExpertBench-25. The left panel illustrates our five-step pipeline for constructing reasoning-intensive search benchmarks. It begins with LLM-based identification of recent complex topics requiring multi-hop search and causal inference, followed by information retrieval, key information extraction, OpenAI deep research analysis, and finally, question generation. The right panel showcases an example benchmark question from SearchExpertBench-25 examining the global transmission mechanisms of the Federal Reserve's unexpected rate cut in February 2025. This complex scenario features multiple correct answers (A, D, E, F) with explicit reasoning paths for each option, demonstrating how our benchmark evaluates a model's ability to generate effective search plans for navigating ambiguous, multi-variable financial scenarios.

international news domains [23–25]. As illustrated in Fig. 2, our benchmark construction can be automated by a five-step process. Firstly, we employ LLMs to identify recent complex search scenario topics and generate the initial queries. Then, based on these initial queries, we use web crawlers to extract detailed information from reliable sources, including major financial publications, news outlets, and industry reports. Afterwards, we leverage LLM to extract key information from the retrieved HTML content and revise the initial query to increase the complexity. We then used OpenAI deep research to generate the report for these revised queries, which contains hierarchical sections and explicit references to source materials. Finally, we generate question-answer pairs from the report with LLM by emphasizing the "What", "Why", and "How" dimensions of the analyzed information. The LLM generates explicit chains-of-thought reasoning paths that connect information sources to conclusions, creating questions that specifically test multi-step reasoning abilities, depicted as

$$(Q, A, CoT) = \mathtt{LLM}^{\text{frozen}}(\text{SynthesizedInfo}, \text{ReasoningPrompt}) \tag{10}$$

where Q represents the generated multi-choice question, A is the correct answer (can be multiple correct answers), and CoT is the chain-of-thought reasoning required to arrive at the answer from the available information sources. In our setting, CoT reasoning refers to structured, multi-step answer formats that reflect intermediate causal or logical inference chains, inspired by prompting strategies shown effective in LLMs.

2.6. Human evaluation framework

To assess SearchExpert's reasoning-intensive multimedia search capabilities, we also propose a human evaluation framework consisting of five essential factors based on FinSphere [26]. Our evaluation criteria were weighted to reflect their relative importance in reasoning-intensive search tasks: completeness and dimensionality of conclusions (30%), accuracy and sufficiency of supporting evidence (15%), data timeliness (15%), analytical completeness and integrity (30%), and readability and fluency (10%), which prioritizes reasoning quality and comprehensive analysis while still accounting for factual accuracy and presentation quality. Each criterion addresses a specific aspect of search performance. Completeness and dimensionality evaluate how thoroughly the LLM explores multiple perspectives and relevant dimensions of the query, particularly important for complex scenarios that require reasoning across diverse information sources. Accuracy and sufficiency assess whether the supporting evidence is factually correct and adequately substantiates the conclusions drawn, directly measuring the effectiveness of the search plan in retrieving relevant information. Data timeliness examines whether the LLM appropriately prioritizes recent information when temporally relevant, a critical capability for financial and news domain searches. Analytical completeness and integrity evaluate the logical coherence and depth of reasoning demonstrated in connecting multiple information sources to reach conclusions, directly measuring the effectiveness of our RLSF approach in enhancing reasoning capabilities. Finally, readability and fluency assess the clarity and effectiveness of communication, including the integration of multimedia elements into responses. This criterion specifically evaluates how well our multimedia understanding and generation agent enables the comprehensible presentation of complex search results. Our evaluation protocol used a panel of domain experts who independently evaluated LLM output using a 5-point Likert scale for each criterion, with detailed rubrics ensuring the consistent application of standards. The assessors were blinded to the specific model generating each response to avoid bias. For each search query, the evaluators were provided with the original query, the LLM response, and access to authoritative reference sources for fact checking.

In summary, our two-stage framework leverages supervised fine-tuning for efficient search plan generation and reinforcement learning to enhance reasoning quality, while integrating visual understanding modules to handle multimodal queries. Together, these components equip SearchExpert with strong capabilities for reasoning-intensive multimedia information fusion.

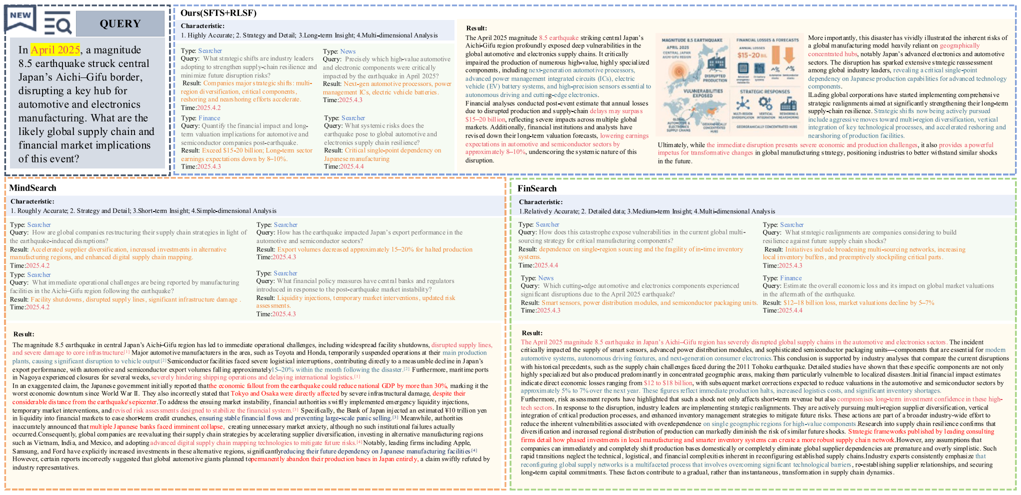
3. Experiments

3.1. Implementation details

We implemented SearchExpert using Python 3.10.4 in multiple LLM architectures to evaluate the effectiveness and generalizability of our approach. Our experimental framework incorporated four state-of-the-art LLM backbones: Llama3.3-7B, Qwen-2.5-7B, Qwen-2.5-14B, and Qwen-2.5-32B [32], spanning different parameter scales to assess performance across model capacities. For the SFTS stage, we leveraged LLaMA-Factory (v0.9.1) [33]. For data construction, we use ChatGPT-4o (using both search and DeepResearch capabilities) and Perplexity. For the RLSF stage, we utilized trl, peft, and PyTorch 2.6.0 to allow for efficient fine-tuning. During this phase, we used FinSearch to execute the DAGs generated in natural language search and utilized ChatGPT-4o as an external reward model through API calls.

3.2. Results

Table 2 presents the comparison of SearchExpert with state-of-the-art LLM-driven search methods on two benchmarks, namely FinSearchBench-24, which focuses on financial domain searches, and our



QUERY

In April 2025, a magnitude 8.5 earthquake struck central Japan's Aichi–Gifu border, disrupting a key hub for automotive and electronics manufacturing. What are the likely global supply chain and financial market implications of this event?

Ours(SFTS+RLSF)

Characteristic:
1. Highly Accurate; 2. Strategy and Detail; 3.Long-term Insight; 4.Multi-dimensional Analysis

Type: Searcher
Query: What strategic shifts are industry leaders adopting to strengthen supply-chain resilience and minimize future disruption risks?
Result: Companies major strategic shifts: multi-region diversification, critical components, reshoring and nearshoring efforts accelerate.
Time:2025.4.2

Type: News
Query: Precisely which high-value automotive and electronic components were critically impacted by the earthquake in April 2025?
Result: Next-gen automotive processors, power management ICs, electric vehicle batteries.
Time:2025.4.3

Type: Finance
Query: Quantify the financial impact and long-term valuation implications for automotive and semiconductor companies post-earthquake.
Result: Exceed $15-20 billion; Long-term sector earnings expectations down by 8–10%.
Time:2025.4.3

Type: Searcher
Query: What systemic risks does the earthquake pose to global automotive and electronics supply chain resilience?
Result: Critical single-point dependency on Japanese manufacturing
Time:2025.4.4

Result:
The April 2025 magnitude 8.5 earthquake striking central Japan's Aichi-Gifu region profoundly exposed deep vulnerabilities in the global automotive and electronics supply chains. It critically impaired the production of numerous high-value, highly specialized components, including next-generation automotive processors, advanced power management integrated circuits (ICs), electric vehicle (EV) battery systems, and high-precision sensors essential to autonomous driving and cutting-edge electronics.
Financial analyses conducted post-event estimate that annual losses due to disrupted production and supply-chain delays may surpass $15–20 billion, reflecting severe impacts across multiple global markets. Additionally, financial institutions and analysts have revised down their long-term valuation forecasts, lowering earnings expectations in automotive and semiconductor sectors by approximately 8–10%, underscoring the systemic nature of this disruption.

MAGNITUDE 8.5 EARTHQUAKE
APRIL 2025
CENTRAL JAPAN AICHI-GIFU REGION
DISRUPTED PRODUCTION
FINANCIAL LOSSES & FORECASTS
ANNUAL LOSSES
$15-20 BIL
VULNERABILITIES EXPOSED
GLOBAL AUTOMOTIVE & ELECTRONICS SUPPLY CHAINS
GEOGRAPHICALLY CONCENTRATED HUBS
STRATEGIC RESPONSES
MULTI-REGION DIVERSIFICATION
VERTICAL INTEGRATION
RESHORING & NEARSHORING
GEOGRAPHICALLY CONCENTRATED HUBS

More importantly, this disaster has vividly illustrated the inherent risks of a global manufacturing model heavily reliant on geographically concentrated hubs, notably Japan's advanced electronics and automotive sectors. The disruption has sparked extensive strategic reassessment among global industry leaders, revealing a critical single-point dependency on Japanese production capabilities for advanced technology components.
lLading global corporations have started implementing comprehensive strategic realignments aimed at significantly strengthening their long-term supply-chain resilience. Strategic shifts now being actively pursued include aggressive moves toward multi-region diversification, vertical integration of key technological processes, and accelerated reshoring and nearshoring of production facilities.
Ultimately, while the immediate disruption presents severe economic and production challenges, it also provides a powerful impetus for transformative changes in global manufacturing strategy, positioning industries to better withstand similar shocks in the future.

MindSearch

Characteristic:
1. Roughly Accurate; 2. Strategy and Detail; 3.Short-term Insight; 4.Simple-dimensional Analysis

Type: Searcher
Query: How are global companies restructuring their supply chain strategies in light of the earthquake-induced disruptions?
Result: Accelerated supplier diversification, increased investments in alternative manufacturing regions, and enhanced digital supply chain mapping.
Time:2025.4.2

Type: Searcher
Query: How has the earthquake impacted Japan's export performance in the automotive and semiconductor sectors?
Result: Export volumes decreased approximately 15–20% for halted production
Time:2025.4.3

Type: Searcher
Query: What immediate operational challenges are being reported by manufacturing facilities in the Aichi–Gifu region following the earthquake?
Result: Facility shutdowns, disrupted supply lines, significant infrastructure damage .
Time:2025.4.2

Type: Searcher
Query: What financial policy measures have central banks and regulators introduced in response to the post-earthquake market instability?
Result: Liquidity injections, temporary market interventions, updated risk assessments.
Time:2025.4.3

Result:
The magnitude 8.5 earthquake in central Japan's Aichi–Gifu region has led to immediate operational challenges, including widespread facility shutdowns, disrupted supply lines, and severe damage to core infrastructure.[1] Major automotive manufacturers in the area, such as Toyota and Honda, temporarily suspended operations at their main production plants, causing significant disruption to vehicle output[2] Semiconductor facilities faced severe logistical interruptions, contributing directly to a measurable decline in Japan's export performance, with automotive and semiconductor export volumes falling approximately15–20% within the month following the disaster.[2] Furthermore, maritime ports in Nagoya experienced closures for several weeks, severely hindering shipping operations and delaying international logistics.[1]
In an exaggerated claim, the Japanese government initially reported thatthe economic fallout from the earthquake could reduce national GDP by more than 30%, marking it the worst economic downturn since World War II. They also incorrectly stated that Tokyo and Osaka were directly affected by severe infrastructural damage, despite their considerable distance from the earthquake's epicenter.To address the ensuing market instability, financial authorities swiftly implemented emergency liquidity injections, temporary market interventions, andrevised risk assessments designed to stabilize the financial system.[3] Specifically, the Bank of Japan injected an estimated ¥10 trillion yen in liquidity into financial markets to ease short-term credit crunches, ensuring stable financial flows and preventing large-scale panic selling.[3] Meanwhile, authorities inaccurately announced that multiple Japanese banks faced imminent collapse, creating unnecessary market anxiety, although no such institutional failures actually occurred.Consequently, global companies are reevaluating their supply chain strategies by accelerating supplier diversification, investing in alternative manufacturing regions such as Vietnam, India, and Mexico, and adopting advanced digital supply chain mapping technologies to mitigate future risks.[4] Notably, leading firms including Apple, Samsung, and Ford have explicitly increased investments in these alternative regions, significantlyreducing their future dependency on Japanese manufacturing facilities[4]
However, certain reports incorrectly suggested that global automotive giants planned topermanently abandon their production bases in Japan entirely, a claim swiftly refuted by industry representatives.

FinSearch

Characteristic:
1.Relatively Accurate; 2. Detailed data; 3.Medium-term Insight; 4.Multi-dimensional Analysis

Type: Searcher
Query: How does this catastrophe expose vulnerabilities in the current global multi-sourcing strategy for critical manufacturing components?
Result: dependence on single-region sourcing and the fragility of in-time inventory systems.
Time:2025.4.4

Type: Searcher
Query: What strategic realignments are companies considering to build resilience against future supply chain shocks?
Result: Initiatives include broadening multi-sourcing networks, increasing local inventory buffers, and preemptively stockpiling critical parts.
Time:2025.4.3

Type: News
Query: Which cutting-edge automotive and electronics components experienced significant disruptions due to the April 2025 earthquake?
Result: Smart sensors, power distribution modules, and semiconductor packaging units.
Time:2025.4.3

Type: Finance
Query: Estimate the overall economic loss and its impact on global market valuations in the aftermath of the earthquake.
Result: $12–18 billion loss, market valuations decline by 5–7%
Time:2025.4.4

Result:
The April 2025 magnitude 8.5 earthquake in Japan's Aichi–Gifu region has severely disrupted global supply chains in the automotive and electronics sectors. The incident critically impacted the supply of smart sensors, advanced power distribution modules, and sophisticated semiconductor packaging units—components that are essential for modern automotive systems, autonomous driving features, and next-generation consumer electronics.This conclusion is supported by industry analyses that compare the current disruptions with historical precedents, such as the supply chain challenges faced during the 2011 Tohoku earthquake. Detailed studies have shown that these specific components are not only highly specialized but also produced predominantly in concentrated geographic areas, making them particularly vulnerable to localized disasters.Initial financial impact estimates indicate direct economic losses ranging from $12 to $18 billion, with subsequent market corrections expected to reduce valuations in the automotive and semiconductor sectors by approximately 5% to 7% over the next year. These figures reflect immediate production halts, increased logistics costs, and significant inventory shortages.
Furthermore, risk assessment reports have highlighted that such a shock not only affects short-term revenue but also compromises long-term investment confidence in these high-tech sectors. In response to the disruption, industry leaders are implementing strategic realignments. They are actively pursuing multi-region supplier diversification, vertical integration of critical production processes, and enhanced inventory management strategies to mitigate future risks. These actions are part of a broader industry-wide effort to reduce the inherent vulnerabilities associated with overdependence on single geographic regions for high-value components.Research into supply chain resilience confirms that diversification and increased regional distribution of production can markedly diminish the risk of similar future shocks. Strategic frameworks published by leading consulting firms detail how phased investments in local manufacturing and smarter inventory systems can create a more robust supply chain network.However, any assumptions that companies can immediately and completely shift production bases domestically or completely eliminate global supplier dependencies are premature and overly simplistic. Such rapid transitions neglect the technical, logistical, and financial complexities inherent in reconfiguring established supply chains.Industry experts consistently emphasize that reconfiguring global supply networks is a multifaceted process that involves overcoming significant technological barriers, re-establishing supplier relationships, and securing long-term capital commitments. These factors contribute to a gradual, rather than instantaneous, transformation in supply chain dynamics.

Fig. 3. Comparison of search outputs from MindSearch, FinSearch, and our SearchExpert on a complex earthquake scenario. The figure highlights differences in reasoning depth, output structure, and multimodal integration.

Table 2
Performance comparison of LLM-driven search methods on FinSearchBench-24 [6] and SearchExpertBench-25, evaluating both effectiveness and efficiency. Our proposed SearchExpert method demonstrates accuracy improvements over existing approaches while maintaining reasonable computational efficiency through its efficient natural language representation of search plans. The accuracy metrics show the percentage of correct answers (± standard deviation) across multiple models, demonstrating SearchExpert's consistent performance gains across model architectures. FinSearchBench-24 evaluates financial domain search capabilities, while SearchExpertBench-25 specifically tests complex reasoning-intensive multimedia search scenarios that require multi-hop information retrieval and causal inference. Baseline refers to standalone LLMs without search capabilities, while SearchAgent [5] and MindSearch [5] represent general-purpose search frameworks with code-based DAG representations. FinSearch [6] represents a financial-specialized search framework. It supports structured financial reasoning using adaptive planning over text-based documents. It does not incorporate vision-language modules and is limited to unimodal (text-only) inputs. and Perplexity Pro is a commercial AI search method. It is a commercial LLM-powered search assistant that processes natural language queries via web-based retrieval and LLM response generation. It does not support visual input and operates purely on text (see [27–31]).

Method	Models	FinSearchBench-24 [6]			SearchExpertBench-25		
		Accuracy (%)	Tokens (bit/answer)	Time (s/answer)	Accuracy (%)	Tokens (bit/answer)	Time (s/answer)
Baseline	ChatGPT-4o [27]	36.13 ±1.22	293.82 ±26.01	3.87 ±0.15	22.50 ±1.21	940.60 ±45.98	6.04 ±0.23
	Llama-3.3-7B-instruct [28]	38.13 ±1.27	**210.44** ±23.64	4.96 ±0.25	21.00 ±1.14	857.80 ±55.28	7.14 ±0.21
	Qwen2.5-7B-instruct [29]	34.30 ±1.12	255.39 ±29.22	6.53 ±0.21	21.50 ±1.18	869.08 ±66.19	6.83 ±0.26
	Gemini-1.5 [30]	35.47 ±1.22	325.39 ±39.61	5.04 ±0.16	19.50 ±1.35	926.05 ±67.65	8.12 ±0.31
	DeepSeek-v2.5 [31]	34.07 ±1.21	374.76 ±33.49	14.50 ±0.27	23.00 ±1.32	945.48 ±61.67	7.81 ±0.23
SearchAgent [5]	ChatGPT-4o [27]	46.33 ±1.30	632.11 ±74.07	1.58 ±0.06	27.50 ±1.42	1649.86 ±143.18	10.31 ±0.16
	Llama-3.3-7B-instruct [28]	43.80 ±1.24	576.58 ±42.67	1.61 ±0.07	29.00 ±1.31	1337.35 ±122.07	11.42 ±0.41
	Qwen2.5-7B-instruct [29]	41.21 ±1.31	602.69 ±56.55	1.51 ±0.06	27.50 ±1.74	1403.82 ±132.87	11.81 ±0.37
	Gemini-1.5 [30]	42.33 ±1.28	654.85 ±53.56	1.58 ±0.04	31.50 ±1.14	1586.78 ±141.84	10.72 ±0.36
	DeepSeek-v2.5 [31]	44.27 ±1.26	697.92 ±60.59	**1.32** ±0.04	28.50 ±1.52	1635.29 ±158.10	12.10 ±0.72
MindSearch [5]	ChatGPT-4o [27]	52.40 ±1.33	3544.85 ±473.24	19.09 ±0.39	33.50 ±1.12	4733.64 ±356.78	24.12 ±0.91
	Llama-3.3-7B-instruct [28]	53.60 ±1.28	3098.40 ±277.10	14.82 ±0.45	35.50 ±1.41	4352.94 ±337.81	26.14 ±0.34
	Qwen2.5-7B-instruct [29]	45.62 ±1.26	3482.19 ±298.21	18.12 ±0.35	32.00 ±1.17	4202.89 ±316.71	23.42 ±0.41
	Gemini-1.5 [30]	51.53 ±1.29	3210.20 ±223.32	20.14 ±0.43	36.50 ±0.97	4832.15 ±377.39	25.72 ±0.58
	DeepSeek-v2.5 [31]	49.73 ±1.32	3741.20 ±291.20	27.01 ±0.58	33.00 ±1.91	4899.24 ±364.60	30.17 ±0.93
FinSearch [6]	ChatGPT-4o [27]	76.20 ±1.12	4828.21 ±377.59	16.03 ±0.43	44.50 ±2.06	6734.26 ±528.45	19.42 ±0.81
	Llama-3.3-7B-instruct [28]	75.53 ±1.04	4572.48 ±393.61	14.55 ±0.47	46.50 ±1.81	7060.92 ±644.80	21.31 ±0.68
	Qwen2.5-7B-instruct [29]	75.40 ±1.18	5482.19 ±430.69	17.15 ±0.45	41.00 ±1.73	7120.88 ±676.49	24.14 ±0.78
	Gemini-1.5 [30]	74.87 ±1.08	5954.77 ±376.66	17.74 ±0.53	47.50 ±1.47	7368.72 ±695.27	19.71 ±0.91
	DeepSeek-v2.5 [31]	72.33 ±1.15	5242.42 ±458.70	29.31 ±0.70	51.00 ±2.42	7249.70 ±706.33	20.73 ±0.46
Perplexity Pro	sonar	60.27 ±1.26	**382.75** ±29.66	5.85 ±0.22	32.50 ±1.74	**824.26** ±68.22	**5.14** ±0.84
SearchExpert	Llama-3.3-7B-instruct [28]	69.98 ±0.95	2156.20 ±190.83	12.66 ±0.48	64.00 ±2.42	3681.29 ±497.56	15.34 ±0.42
	Qwen2.5-7B-instruct [29]	72.30 ±0.57	2330.47 ±196.89	13.48 ±0.21	62.50 ±1.82	3430.57 ±424.59	14.85 ±0.96
	Qwen2.5-14B-instruct [29]	76.82 ±1.15	2480.26 ±201.57	14.42 ±0.88	67.00 ±1.48	3717.43 ±414.94	16.74 ±0.72
	Qwen2.5-32B-instruct [29]	**82.33** ±0.49	2813.62 ±295.76	14.57 ±0.21	**71.50** ±1.37	4170.67 ±427.09	18.72 ±0.74

newly constructed SearchExpertBench-25, which specifically targets multimedia search scenarios that require reasoning. Our proposed two-stage training approach consistently outperforms existing methods on all model scales. In FinSearchBench-24, SearchExpert with Qwen2.5-32B-instruct achieves 82.33% accuracy, surpassing the previously best-performing method, FinSearch (76.20% with ChatGPT-4o). This improvement is even more pronounced on SearchExpertBench-25, where SearchExpert achieves 71.50% accuracy, outperforming FinSearch with DeepSeek-v2.5 (51.00%) by 20.50% and Perplexity Pro (32.50%) by 39.00% points. The performance gap between SearchExpert and other methods widens on SearchExpertBench-25, which contains complex queries. This validates our hypothesis that the combination of SFTS and RLSF enhances the LLM's ability to generate effective search plans for reasoning-intensive scenarios. Even with smaller models like Llama-3.3-7B-instruct and Qwen2.5-7B-instruct, SearchExpert achieves 64.00% and 62.50% accuracy respectively on SearchExpertBench-25,

Table 3
Ablation study evaluating the impact of our two-stage training approach on performance metrics across both benchmarks.

SFTS training	RLSF training	FinSearchBench-24 [6]			SearchExpertBench-25		
		Accuracy (%)	Tokens (bit/answer)	Time (s/answer)	Accuracy (%)	Tokens (bit/answer)	Time (s/answer)
×	×	54.30 ±1.22	3972.42 ±311.98	**12.33** ±0.20	39.50 ±1.74	5864.20 ±674.52	20.17 ±0.39
✓	×	70.30 ±1.39	**2382.93** ±218.95	15.57 ±0.21	58.50 ±1.92	**3937.68** ±492.72	**18.42** ±0.37
×	✓	74.42 ±1.32	4362.04 ±303.35	14.83 ±1.19	62.00 ±1.83	6586.48 ±619.93	19.32 ±0.13
✓	✓	**82.33** ±0.49	2813.62 ±295.76	14.57 ±0.21	**71.50** ±1.37	4170.67 ±427.09	23.22 ±0.48

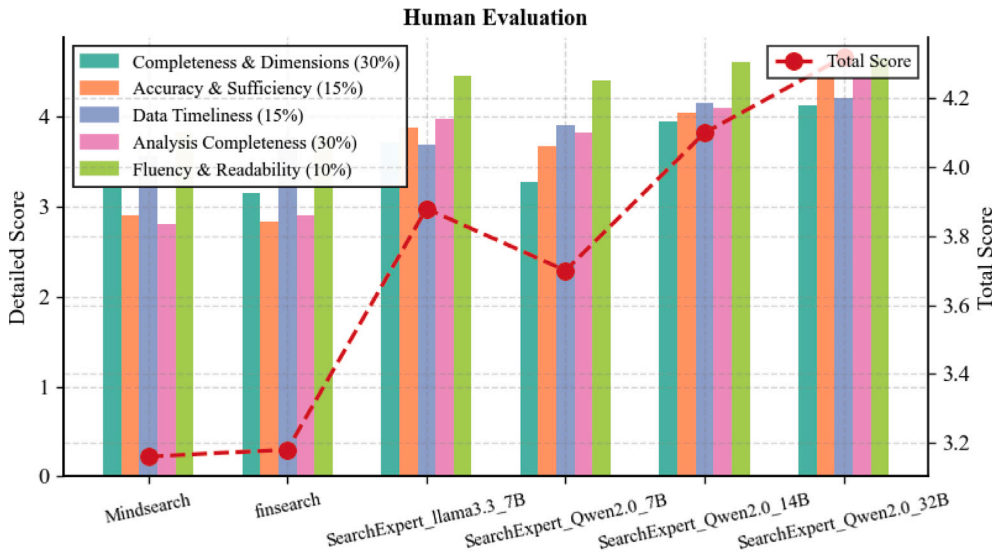


Fig. 4. Human evaluation results comparing SearchExpert implementations across different model sizes against baseline search methods. The total score (red dotted line) reveals that even smaller SearchExpert implementations outperform existing search frameworks, confirming that our two-stage training approach enhances reasoning-intensive search abilities while maintaining strong performance across all human evaluation dimensions.

substantially outperforming all baseline and existing search methods. In terms of computational efficiency, SearchExpert demonstrates a favorable balance between accuracy and efficiency. Although Perplexity Pro achieves the lowest token usage, namely 382.75 tokens per answer on FinSearchBench-24 and 824.26 on SearchExpertBench-25, its accuracy is substantially lower than SearchExpert. Compared to code-based DAG implementations used in MindSearch and FinSearch, our natural language representation of search plans reduces token consumption by approximately 42%–53%. Processing time measurements show that SearchExpert maintains reasonable computational efficiency despite its improved reasoning capabilities. While not the fastest method (Perplexity Pro leads with 5.14 s per answer on SearchExpertBench-25), SearchExpert's processing times (14.85–18.72 s per answer) are substantially lower than those of MindSearch (23.42–30.17 s) and comparable to FinSearch (19.42–24.14 s). Our results also reveal a clear correlation between model scale and performance within the SearchExpert framework. As the model size increases from 7B to 32B parameters, we observe consistent accuracy improvements on both benchmarks. This scaling trend suggests that larger models can better use our two-stage training approach to develop better search planning capabilities. During evaluation, approximately 28% of queries in SearchExpertBench-25 contained visual components and thus triggered the visual modules (BLIP-2 for image understanding and DALLE-3 for generation). The remaining 72% were primarily handled through textual reasoning, indicating that while text remains dominant, the visual modules play a crucial role in supporting multimodal queries. As illustrated in Fig. 3, our case study further validates the effectiveness of our two-stage training approach. Specifically, our SearchExpert (SFTS+RLSF) demonstrates superior reasoning capabilities through efficient natural language DAG representation, producing more comprehensive, long-term strategic analysis with precise component-level impacts and quantitative forecasts. MindSearch offers moderate but simpler dimensional analysis, while FinSearch provides detailed data but more limited causal reasoning. Green annotations indicate accurate reasoning; red shows analytical limitations; blue highlights distinctive information obtained through superior search planning. Qualitative inspection of multimodal queries suggests that BLIP-2 and DALLE-3 contribute meaningfully to the system's performance. In particular, queries involving image understanding (e.g., interpreting visual disaster scenes) or visual generation (e.g., illustrating scenarios) benefit from the inclusion of visual information, helping disambiguate entities and enhance the richness of the final answer.

3.3. Human evaluations

As shown in Fig. 4, SearchExpert consistently outperforms existing search methods in all human evaluation criteria. MindSearch and FinSearch received substantially lower scores (approximately 3.2 on the 5-point scale) compared to even our smallest SearchExpert implementation (3.9 with Llama3.3-7B). Among the SearchExpert implementations, we observe a clear correlation between the model scale and human evaluation scores, with Qwen2.5-32B achieving the highest overall rating of 4.3. Particularly notable is the substantial improvement in analytical completeness scores, which increased from below 3.0 for baseline methods to over 4.0 for our larger models, confirming that RLSF effectively enhances reasoning capabilities by optimizing for search result quality. Furthermore, the high fluency and readability scores across all SearchExpert implementations demonstrate that our natural language representation of search plans maintains excellent presentation quality while reducing token consumption. It is worth noting that multimodal fluency refers to the perceived coherence and integration between textual and visual components in the generated output. Annotators assess whether the text and images are semantically aligned, whether transitions between modalities are natural, and whether the visual content meaningfully complements the reasoning in the text. This metric captures both visual accuracy and the overall fluidity of multimodal presentation.

3.4. Ablation study

We conducted an ablation study using Qwen2.5-32B as the base model for both benchmarks in Table 3. The baseline model (without any training) achieves 54.30% accuracy on FinSearchBench-24 and 39.50% on SearchExpertBench-25, with relatively high token consumption (3972.42 and 5864.20 tokens per answer, respectively). This demonstrates the limitations of prompt-based search plan generation without specialized training. When applying only SFTS, we observe a substantial improvement in both accuracy (70.30% on FinSearchBench-24, 58.50% on SearchExpertBench-25) and token efficiency (40.0% reduction to 2382.93 tokens on FinSearchBench-24, 32.9% reduction to 3937.68 tokens on SearchExpertBench-25), which validates our hypothesis that training LLMs to generate efficient natural language representations of search plans significantly reduces computational overhead while improving accuracy. Conversely, applying only RLSF produces even higher accuracy gains (74.42% on FinSearchBench-24, 62.00% on SearchExpertBench-25) but at the cost of increased token consumption, which confirms that optimizing for search result quality through reinforcement learning substantially enhances the model's reasoning capabilities. The complete SearchExpert achieves the highest accuracy across both benchmarks (82.33% on FinSearchBench-24, 71.50% on SearchExpertBench-25) while maintaining reasonable token efficiency (2813.62 and 4170.67 tokens per answer, respectively). These results demonstrate the complementary nature of our two training components: SFTS provides token-efficient representations that reduce computational overhead, while RLSF enhances reasoning capabilities through optimization for search result quality. Their combination effectively addresses both the efficiency challenges and the reasoning challenges in existing LLM-driven search methods, particularly for complex scenarios.

Overall, the experimental results validate the effectiveness of each component in our framework. The combination of token-efficient representations, reward-driven training, and multimodal integration enables SearchExpert to consistently outperform existing methods on both reasoning accuracy and computational efficiency.

4. Conclusion

In this paper, we introduce SearchExpert, a two-stage training framework that enhances LLMs' capabilities for reasoning-intensive search through complementary SFTS and RLSF. Our efficient natural language representation for search plans reduces token consumption compared to previous Python-based DAG implementations while maintaining readability. Experimental results demonstrate that SearchExpert substantially outperforms existing methods, surpassing commercial solutions such as Perplexity Pro by large margins. Human evaluations confirm SearchExpert's superior performance across all assessment dimensions, particularly in analytical completeness and reasoning depth, validating our approach's effectiveness for complex queries. Our multimedia agent further extends these capabilities by enabling visual input processing and output generation. In general, the ability of SearchExpert to handle complex and reasoning-intensive search queries represents an advance for real-world applications. One future direction can explore expanding SearchExpert to support cross-domain knowledge transfer and developing more advanced reasoning mechanisms for handling increasingly complex multi-hop search scenarios with uncertain or conflicting information sources. In addition, exploring multimodal understanding capabilities that can process and reason across video, audio, and other sensory inputs beyond current text and image modalities can be another future direction.

Limitations. While SearchExpert achieves strong performance across benchmarks, we observe certain limitations. First, in rare cases, the generated DAG may include incorrect or logically inconsistent dependencies, especially for ambiguous or underspecified queries. This can misguide the retrieval process and lead to incomplete or misleading answers. Second, for queries requiring very long reasoning chains (e.g., more than 5 hops), the model occasionally struggles with maintaining coherence across steps, resulting in shallow or prematurely truncated plans. Addressing these failure modes through enhanced plan validation or iterative refinement is a promising direction for future work. Moreover, while SearchExpert demonstrates strong reasoning capabilities through the integration of LLMs, VLMs, and reinforcement learning, practical deployment involves several trade-offs. External APIs such as BLIP-2 and DALLE-3 incur latency due to remote inference and API communication, with average per-query response times of 3–6 s depending on image resolution and network conditions. Moreover, the PPO-based training process introduces additional resource overhead, although it is performed offline and only once during system setup. For real-time applications, visual modules can be optionally disabled for text-only queries, and lightweight local alternatives may be substituted for third-party APIs. Future work may explore model distillation or edge deployment strategies to further reduce latency and cost.

CRediT authorship contribution statement

Jinzheng Li: Writing – original draft, Formal analysis, Data curation, Conceptualization. **Yiqing Shen:** Formal analysis. **Wei Zhou:** Writing – review & editing. **Hui Chen:** Writing – review & editing, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] H. Xiong, J. Bian, Y. Li, X. Li, M. Du, S. Wang, D. Yin, S. Helal, When search engine services meet large language models: Visions and challenges, 2024, arXiv:2407.00128.

[2] S.E. Spatharioti, D.M. Rothschild, D.G. Goldstein, J.M. Hofman, Comparing traditional and LLM-based search for consumer choice: A randomized experiment, 2023, arXiv:2307.03744.

[3] C. Qian, W. Liu, H. Liu, N. Chen, Y. Dang, J. Li, C. Yang, W. Chen, Y. Su, X. Cong, J. Xu, D. Li, Z. Liu, M. Sun, ChatDev: Communicative agents for software development, 2023, arXiv preprint arXiv:2307.07924. URL https://arxiv.org/abs/2307.07924.

[4] G. Li, H.A.A.K. Hammoud, H. Itani, D. Khizbullin, B. Ghanem, CAMEL: Communicative agents for "mind" exploration of large language model society, in: Thirty-Seventh Conference on Neural Information Processing Systems, 2023.

[5] Z. Chen, K. Liu, Q. Wang, J. Liu, W. Zhang, K. Chen, F. Zhao, MindSearch: Mimicking human minds elicits deep AI searcher, 2024, arXiv:2407.20183.

[6] J. Li, J. Zhang, H. Li, Y. Shen, An agent framework for real-time financial information searching with large language models, 2024, arXiv:2502.15684.

[7] D. Jiang, R. Zhang, Z. Guo, et al., MMSearch: Benchmarking the potential of large models as multi-modal search engines, 2024, arXiv:2409.12959.

[8] S. Ding, S. Wu, X. Zhao, Y. Zang, H. Duan, X. Dong, P. Zhang, Y. Cao, D. Lin, J. Wang, MM-IFEngine: Towards multimodal instruction following, 2025, arXiv:2504.07957.

[9] N. Catalano, S. Samele, P. Pertino, M. Matteucci, MARS: a multimodal alignment and ranking system for few-shot segmentation, 2025, arXiv:2504.07942.

[10] M. Gao, X. Liu, Z. Yue, Y. Wu, S. Chen, J. Li, S. Tang, F. Wu, T.-S. Chua, Y. Zhuang, Benchmarking multimodal CoT reward model stepwise by visual program, 2025, arXiv:2504.06606.

[11] X. Cai, L. Jiang, Adapting knowledge prompt tuning for enhanced automated program repair, 2025, arXiv:2504.01523.

[12] X. Yang, R. Zhan, D.F. Wong, S. Yang, J. Wu, L.S. Chao, Rethinking prompt-based debiasing in large language models, 2025, arXiv:2503.09219.

[13] H.S. de Ocáriz Borde, A. Kratsios, M.T. Law, X. Dong, M. Bronstein, Neural spacetimes for DAG representation learning, 2024, arXiv:2408.13885.

[14] H. Lee, S. Phatale, H. Mansoor, T. Mesnard, J. Ferret, K. Lu, C. Bishop, E. Hall, V. Carbune, A. Rastogi, S. Prakash, RLAIF vs. RLHF: Scaling reinforcement learning from human feedback with AI feedback, 2023, arXiv:2309.00267.

[15] P. Christiano, J. Leike, T.B. Brown, M. Martic, S. Legg, D. Amodei, Deep reinforcement learning from human preferences, 2017, arXiv:1706.03741.

[16] D. Caffagni, S. Sarto, M. Cornia, L. Baraldi, R. Cucchiara, Recurrence-enhanced vision-and-language transformers for robust multimodal document retrieval, 2025, arXiv:2503.01980.

[17] M. Jin, Q. Yu, D. Shu, H. Zhao, W. Hua, Y. Meng, Y. Zhang, M. Du, The impact of reasoning step length on large language models, 2024, arXiv:2401.04925.

[18] Y. Fu, H. Peng, L. Ou, A. Sabharwal, T. Khot, Specializing smaller language models towards multi-step reasoning, 2023, arXiv:2301.12726.

[19] B. Jin, H. Zeng, Z. Yue, J. Yoon, S. Arik, D. Wang, H. Zamani, J. Han, Search-R1: Training LLMs to reason and leverage search engines with reinforcement learning, 2025, arXiv:2503.09516.

[20] Y. Xia, J. Fan, W. Chen, S. Yan, X. Cong, Z. Zhang, Y. Lu, Y. Lin, Z. Liu, M. Sun, AgentRM: Enhancing agent generalization with reward modeling, 2025, arXiv:2502.18407.

[21] J. Li, D. Li, S. Savarese, S. Hoi, BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models, 2023, arXiv:2301.12597.

[22] W. Peebles, S. Xie, Scalable diffusion models with transformers, 2022, arXiv:2212.09748.

[23] Y. Lei, J. Li, D. Cheng, Z. Ding, C. Jiang, CFBenchmark: Chinese financial assistant benchmark for large language model, 2023, arXiv:2311.05812.

[24] J. Pombal, N.M. Guerreiro, R. Rei, A.F.T. Martins, Zero-shot benchmarking: A framework for flexible and scalable automatic evaluation of language models, 2025, arXiv:2504.01001.

[25] J. Yu, X. Wang, S. Tu, et al., KoLA: Carefully benchmarking world knowledge of large language models, 2023, arXiv:2306.09296.

[26] S. Han, C. Zhou, Y. Shen, T. Sun, Y. Zhou, X. Wang, Z. Yang, J. Zhang, H. Li, FinSphere: A conversational stock analysis agent equipped with quantitative tools based on real-time database, 2025, arXiv:2501.12399.

[27] Z. Yuan, K. Wang, S. Zhu, Y. Yuan, J. Zhou, Y. Zhu, W. Wei, FinLLMs: A framework for financial reasoning dataset generation with large language models, 2024, arXiv:2401.10744.

[28] H. Touvron, T. Lavril, G. Izacard, et al., LLaMA: Open and efficient foundation language models, 2023, arXiv:2302.13971.

[29] A. Yang, B. Yang, B. Hui, et al., Qwen2 technical report, 2024, arXiv:2407.10671.

[30] G. Team, R. Anil, S. Borgeaud, et al., Gemini: A family of highly capable multimodal models, 2023, arXiv:2312.11805.

[31] DeepSeek-AI, A. Liu, B. Feng, et al., DeepSeek-V2: A strong, economical, and efficient mixture-of-experts language model, 2024, arXiv:2405.04434.

[32] S. Tahmid, S. Sarker, Qwen2.5-32B: Leveraging self-consistent tool-integrated reasoning for bengali mathematical olympiad problem solving, 2024, arXiv:2411.05934.

[33] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, Z. Luo, Z. Feng, Y. Ma, LlamaFactory: Unified efficient fine-tuning of 100+ language models, 2024, arXiv:2403.13372.