Towards Self-Evolving Agentic Literature Retrieval

Yuwen Du [*1 2] **Tian Jin** [*1 2] **Jing Kang** [1 2] **Xianghe Pang** [1 2] **Jingyi Chai** [1 2] **Tingjia Miao** [1 2] **Fenyi Liu** [1 2] **WenHao Wang** [3] **Sikai Yao** [2] **Yuzhi Zhang** [2] **Siheng Chen** [1 2]

[1]Shanghai Jiao Tong University [2]SciLand [3]Zhejiang University

Abstract

As large language models reshape scientific research, literature retrieval faces a twofold challenge: ensuring source authenticity while maintaining a deep comprehension of academic search intents. While reliable, traditional keyword-centric search fails to capture complex research intents. Frontier LLMs can handle complex research intents, but their high cost and tendency to hallucinate remain key limitations. Here we introduce PaSaMaster, a self-evolving agentic literature retrieval system that produces relevance-scored paper rankings with evidence-grounded recommendations through iterative intent analysis, retrieval, and ranking. It is built on three key designs. First, it transforms literature retrieval from a one shot query–document matching problem into a search process that evolves over time, using ranked evidence to reveal gaps, refine intents, and guide follow-up searches. Second, it prevents hallucinated sources by treating retrieval as intent–paper relevance ranking rather than generation. Finally, PaSaMaster improves cost efficiency by separating planning from retrieval: a frontier LLM is used only for intent understanding, while large scale retrieval and relevance scoring are delegated to customized corpora and lightweight models. Evaluated on the PaSaMaster Benchmark across 38 scientific disciplines, our system exposes the severe inaccuracy and incompleteness of traditional keyword retrieval (improving F1-score by 16.5×) and the unreliability of generative LLMs (which exhibit hallucination rates up to 32.66%). Remarkably, PaSaMaster outperforms GPT-5.2 by 37.8% at a mere 1% of the computational cost while ensuring zero source hallucination: https://github.com/sjtu-sai-agents/PaSaMaster.

[*]Equal contribution

1. Introduction

Scientific literature retrieval is the axiomatic starting point of all scientific inquiry (Gusenbauer & Haddaway, 2021). Before formulating hypotheses, designing experiments, or building new theories, researchers must fundamentally navigate the vast and ever-expanding corpus of existing knowledge (Fortunato et al., 2018). However, the volume of scientific publications has grown exponentially over recent decades, decisively overwhelming the fixed cognitive bandwidth of individual researchers (Bornmann & Mutz, 2015). This severe information overload has driven an inevitable reliance on artificial intelligence to automate and accelerate knowledge discovery (Wang et al., 2023). More importantly, modern literature search is rarely a simple keyword lookup. Researchers often express complex academic intents involving technical constraints, application contexts, and implicit background knowledge (White & Roth, 2009; Gusenbauer & Haddaway, 2021; Ajith et al., 2024). As large language models reshape scientific research workflows, literature retrieval therefore faces a new central challenge: how to deeply understand complex research intents while ensuring that every returned source is real and verifiable (Zhang et al., 2025b; Ajith et al., 2024).

Yet, existing methods still fail to jointly satisfy these two requirements. They either preserve source authenticity at the cost of shallow intent understanding (Google, n.d.; Asai et al., 2024; Zhang et al., 2025a), or improve semantic comprehension while sacrificing factual reliability or scalability (DeepSeek-AI et al., 2025; Team et al., 2026; MiniMax et al., 2025; Team et al., 2025; Google DeepMind, 2026; OpenAI, 2026). This creates a persistent trade-off between verifiable but limited retrieval and more intelligent but less trustworthy literature discovery.

This trade-off becomes clearer when viewed through the evolution of literature retrieval paradigms (Table 1). *Level 0 (Lexical Retrieval)* (Google, n.d.; National Center for Biotechnology Information, 1996) guarantees source authenticity through indexed databases, but reduces complex research intents to rigid keywords, causing severe intent compression. *Level 1 (Semantic Retrieval)* (Asai et al.,

Table 1. **The Five Paradigms of Scientific Literature Discovery.** Existing paradigms each improve one aspect of literature retrieval, but fail to jointly achieve adaptive intent understanding, hallucination-free evidence grounding, and cost-efficient scaling. PaSaMaster resolves these limitations through agentic self-evolving, evidence-grounded retrieval.

Paradigm Level	Representative Systems	Intent Adaptivity	Source Reliability	Cost Efficiency
Level 0: Lexical Retrieval	Google Scholar (Google, n.d.), PubMed (National Center for Biotechnology Information, 1996)	Keyword-based; severe intent compression	Verified indexed papers	Efficient but semantically shallow
Level 1: Semantic Retrieval	OpenScholar (Asai et al., 2024), Bohrium Navigator (Zhang et al., 2025a)	Passive embedding matching; limited intent compression	Verified indexed papers	Efficient but semantically shallow
Level 2: Generative LLMs	GPT-5.2 (OpenAI, 2026), Gemini 3.1 Pro (Google DeepMind, 2026), DeepSeek (DeepSeek-AI et al., 2025)	Strong natural-language understanding	Prone to hallucinated papers	Expensive to deploy at scale
Level 3: Fixed-Pipeline Agentic Retrieval	Google Scholar Labs (Google, 2025), PaSa (He et al., 2025)	User intent fixed at the outset; no cognition update during retrieval	Verified indexed papers	Cost-controlled, but constrained by fixed intent interpretation
Level 4: Self-Evolving Agentic Retrieval	**PaSaMaster (Ours)**	**Iteratively refines intent using ranked evidence**	**Verified indexed papers**	**Cost-efficient planning–retrieval separation**

2024; Zhang et al., 2025a) improves over exact keyword matching by using embedding-based similarity, but still treats retrieval as passive query–document matching and lacks the ability to actively clarify, decompose, or refine complex intents. *Level 2 (Generative LLMs)* (DeepSeek-AI et al., 2025; Team et al., 2026; MiniMax et al., 2025; Team et al., 2025; Google DeepMind, 2026; OpenAI, 2026) offers stronger intent comprehension, yet its probabilistic generation introduces fabricated papers, undermining the factual trust required for scientific inquiry. *Level 3 (Fixed-Pipeline Agentic Retrieval)* (Google, 2025; He et al., 2025) mitigates source hallucination by grounding LLM agents in verifiable retrieval tools. However, these systems typically follow a predefined retrieve–read–answer pipeline, where the user intent is fixed at the outset and retrieval is executed without iterative cognitive updates. As a result, they cannot iteratively refine user intent from ranked evidence, limiting their understanding of complex research needs.

These limitations motivate *Level 4 (Self-Evolving Agentic Retrieval)*, represented by **PaSaMaster**. PaSaMaster is an agentic self-evolving literature retrieval system that produces relevance-scored paper rankings with evidence-grounded recommendations through iterative intent analysis, retrieval, ranking, and refinement. Rather than treating literature search as one-shot query–document matching problem, PaSaMaster formulates scientific literature discovery as a self-evolving intent–paper relevance ranking process. This design enables the system to align with complex research intents while ensuring that every returned source is real, verifiable, and grounded in customized corpora.

PaSaMaster is built on three key designs. First, self-evolving retrieval: it transforms literature retrieval from one-shot query–document matching into an adaptive search process that evolves over time, where retrieved and ranked evidence is used to identify coverage gaps, refine the research intent, and guide subsequent retrieval rounds. Second, hallucination-free ranking: it treats literature discovery as intent–paper relevance ranking rather than generation, ensuring that all recommended papers come from verified corpora and are grounded in original paper evidence. Third, cost-efficient separation: it uses frontier LLMs only for intent understanding and refinement, while delegating large-scale retrieval and relevance scoring to customized scientific corpora and lightweight models. Together, these designs enable PaSaMaster to align with complex research intents while maintaining source verifiability and scalable efficiency.

To evaluate retrieval capability on complex natural-language literature search problems, we introduce **PaSaMaster-Bench**, the first multidisciplinary literature retrieval benchmark designed for complex search intents. Unlike conventional retrieval benchmarks built around short keyword queries, PaSaMaster-Bench focuses on highly specific, multi-constrained natural language search intents that require systems to search, verify, and rank all papers satisfying explicit criteria. The benchmark contains **244 expert-curated tasks** spanning **38 scientific disciplines**, with queries, constraints, target paper lists, and evaluation checklists annotated and verified by human domain experts.

We evaluate PaSaMaster on the PaSaMaster-Bench. The results reveal the severe inaccuracy and incompleteness of traditional keyword retrieval, with PaSaMaster improving F1-score by **16.5**$\times$. They also expose the unreliability of generative LLMs, which exhibit hallucination rates up to **32.66%**. Remarkably, PaSaMaster outperforms GPT-5.2 by **37.8%** while using only **1%** of its computational cost, and maintains zero source hallucination. These results demonstrate that self-evolving, evidence-grounded relevance ranking provides a scalable and trustworthy foundation for AI-assisted scientific literature discovery.

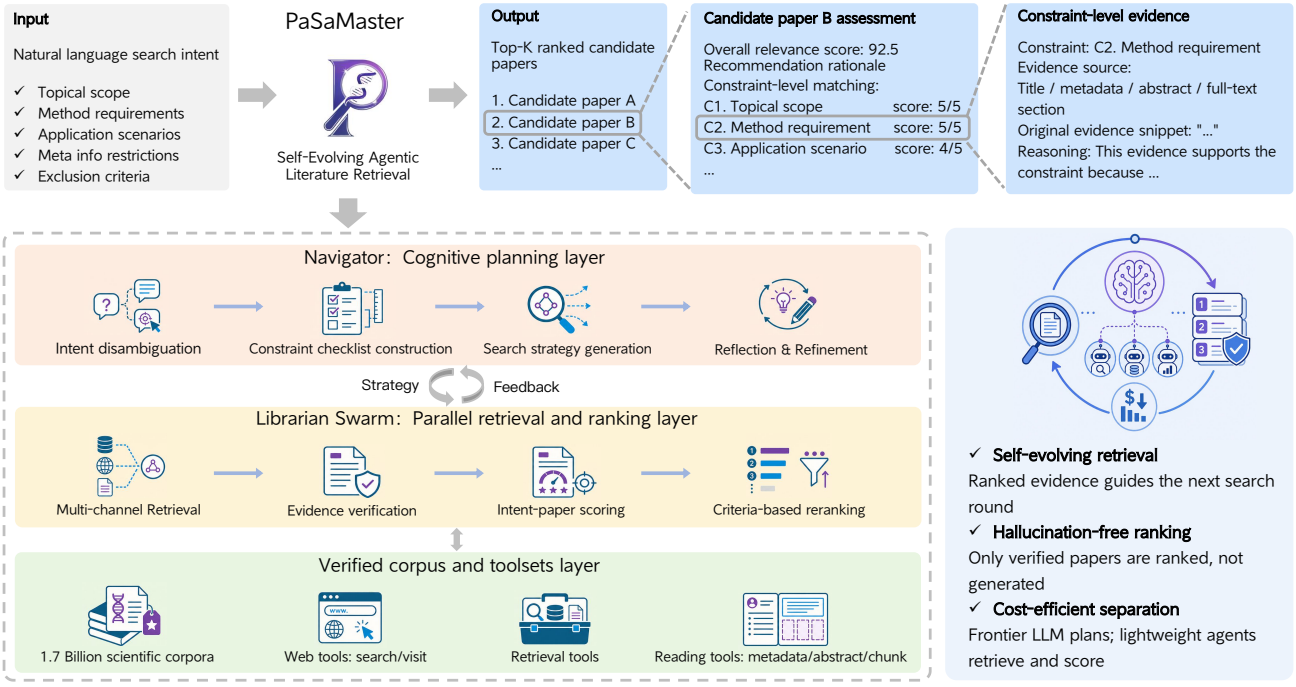

Figure 1. Overview of PaSaMaster. PaSaMaster is a self-evolving agentic literature retrieval system that separates intent-aware planning from evidence-grounded retrieval and ranking. Given a natural-language query, the Navigator disambiguates complex academic intents, generates adaptive search strategies, and refines them through feedback. The Librarian Swarm executes parallel retrieval, evidence verification, intent–paper scoring, and criteria-based reranking over customized corpora and tools. The system outputs hallucination-free, evidence-ranked paper recommendations with scores, supporting evidence, and checkpoint-level rationales.

2. Methodology

PaSaMaster is an agentic self-evolving literature retrieval system that maps a complex natural-language search intent q to a ranked, evidence-grounded paper set $\mathcal{P} = [p_1, p_2, \ldots, p_k]$ without human intervention. Its design follows three principles that directly address the limitations of existing literature retrieval paradigms: *self-evolving retrieval*, *hallucination-free intent–paper relevance ranking*, and *cost-efficient planning–retrieval separation*. Rather than generating paper lists from parametric memory, PaSaMaster retrieves real papers from customized scientific corpora, verifies their relevance using original evidence, and iteratively refines the search intent based on ranked retrieval results.

Formally, PaSaMaster operates over a customized scientific corpus $\mathcal{D}$ and an agent-accessible operator toolset $\mathcal{T}$. Given query q, a Navigator first produces a retrieval strategy S and a query-specific verification checklist $C = [c_1, c_2, \ldots, c_m]$, where each checkpoint c_j encodes one concrete requirement that a relevant paper must satisfy. A swarm of Librarian agents then retrieves candidate papers, verifies them against C, and reranks them into the final output:

$$\langle S,\, C\rangle = \textsc{Plan}(q, \pi_{\text{Nav}}), \tag{1}$$

$$\mathcal{P}_{\text{init}} = \textsc{Retrieve}\Big(S; \mathcal{D}, \mathcal{T}, \{\pi_{\text{Lib}}^{(i)}\}_{i=1}^{N}\Big)\,, \tag{2}$$

$$\mathcal{P}_{\text{scored}} = \textsc{Verify}\Big(\mathcal{P}_{\text{init}}, C; \mathcal{D}, \mathcal{T}, \{\pi_{\text{Lib}}^{(i)}\}_{i=1}^{N}\Big)\,, \tag{3}$$

$$\mathcal{P} = \textsc{Rerank}(\mathcal{P}_{\text{scored}})\,, \tag{4}$$

where π_{Nav} denotes the Navigator policy and $\{\pi_{\text{Lib}}^{(i)}\}_{i=1}^{N}$ denotes N parallel Librarian agents. The Navigator is responsible for intent understanding and strategic planning, while the Librarian swarm executes retrieval, evidence verification, and ranking over $\mathcal{D}$ through $\mathcal{T}$.

2.1. Self-Evolving Retrieval from Ranked Evidence

The first core design of PaSaMaster is to transform literature retrieval from one-shot query–document matching into a self-evolving search process. Existing retrieval systems typically fix their interpretation of the user query at the beginning and execute retrieval under this static understanding. PaSaMaster instead treats retrieval as an iterative process in which ranked evidence is used to update the system's understanding of the research intent.

The process is coordinated by the Navigator agent. Given the initial query q, the Navigator first analyzes the user's

research intent and generates two outputs: a retrieval strategy S, specifying what should be searched, and a verification checklist C, specifying how candidate papers should be judged. The Librarian swarm then retrieves and scores candidate papers. After each retrieval round, the Navigator inspects the ranked results, identifies missing coverage, ambiguous constraints, or under-explored directions, and refines the strategy and checklist for the next round:

$$\langle S^{(t+1)}, C^{(t+1)} \rangle = \textsc{Reflect}\Big(q, S^{(t)}, C^{(t)}, \mathcal{P}^{(t)}_{\text{scored}}\Big), \quad (5)$$

where t indexes the retrieval round. This closed-loop mechanism allows PaSaMaster to progressively improve its interpretation of complex research intents, rather than relying on a fixed query representation determined before retrieval begins.

2.2. Hallucination-Free Intent–Paper Relevance Ranking

The second core design is to prevent hallucinated sources by formulating literature discovery as intent–paper relevance ranking rather than generation. PaSaMaster never asks an LLM to synthesize citations or paper lists directly from parametric memory. Instead, every candidate paper must be retrieved from a verified scientific corpus $\mathcal{D}$, and every relevance judgment must be grounded in traceable evidence from the original paper.

To support verifiable retrieval and evidence grounding, PaSaMaster restructures over 160 million papers into a three-tier agent-native repository:

$$\mathcal{D} = \{\mathcal{D}^{\text{meta}}, \mathcal{D}^{\text{abs}}, \mathcal{D}^{\text{chunk}}\}, \quad (6)$$

where $\mathcal{D}^{\text{meta}}$ stores structured metadata, $\mathcal{D}^{\text{abs}}$ stores abstract-level representations for coarse semantic filtering, and $\mathcal{D}^{\text{chunk}}$ stores passage-level evidence chunks segmented from full texts. For each candidate paper p and checklist item c_j, the Evidence Chunk Locator retrieves supporting passages:

$$\mathcal{E}_p(c_j) = \underset{e \in \mathcal{D}^{\text{chunk}}_p}{\arg\,\text{top-k}} \cos(\phi(c_j), \phi(e)), \quad (7)$$

where $\phi(\cdot)$ is the shared text encoder and $\mathcal{D}^{\text{chunk}}_p$ denotes the chunk set of paper p. This design binds each relevance judgment to explicit textual evidence instead of relying on unsupported model inference.

Each candidate paper is then evaluated by a trained Scorer model. For every checkpoint c_j, the Scorer outputs a satisfaction score $s_j \in \{1, 2, 3, 4, 5\}$ and an evidence-grounded rationale. The checkpoint scores are averaged into a criterion-level relevance signal:

$$\bar{s}(p) = \frac{1}{m} \sum_{j=1}^{m} s_j. \quad (8)$$

To incorporate holistic confidence, PaSaMaster also extracts the Scorer model's calibrated output probability $\rho \in (0, 1)$ for its overall relevance judgment. The final relevance score is:

$$\mathcal{S}(p) = \frac{\bar{s}(p) + \rho}{6}, \quad (9)$$

where the denominator normalizes the maximum possible value of $\bar{s}(p) + \rho$. The top candidates are then passed to a listwise reranker for global cross-paper comparison. The final result $\mathcal{P}$ is therefore a relevance-ranked list of real papers, with each recommendation traceable to paper-level evidence.

2.3. Cost-Efficient Planning–Retrieval Separation

The third core design is planning–retrieval separation, which improves scalability by using frontier LLMs only where they are most valuable. Frontier LLMs are effective for understanding, decomposing, and refining complex research intents, but using them for every retrieval, reading, and ranking operation would be unnecessarily expensive. PaSaMaster therefore assigns high-level reasoning to the Navigator and delegates large-scale retrieval and relevance scoring to customized corpora, and lightweight parallel Librarian agents.

The operator toolset is divided into retrieval and reading tools:

$$\mathcal{T} = \mathcal{T}^{\text{ret}} \cup \mathcal{T}^{\text{read}}. \quad (10)$$

The retrieval tools $\mathcal{T}^{\text{ret}}$ construct a broad candidate pool through complementary retrieval channels, including Semantic Direct Retrieval, Citation Network Expansion, and Web-to-Repository Verification:

$$\mathcal{C}_{\text{init}} = \bigcup_{o \in \mathcal{T}^{\text{ret}}} \textsc{Retrieve}_o(S, \mathcal{D}). \quad (11)$$

Semantic Direct Retrieval provides high-precision semantic candidates, Citation Network Expansion follows citation links to surface structurally related papers, and Web-to-Repository Verification maps external web findings back to verified repository entries. The reading tools $\mathcal{T}^{\text{read}}$ then support efficient metadata lookup, abstract reading, and evidence-chunk localization, avoiding expensive full-document reading and substantially reducing computational cost.

Finally, to equip each Librarian agent with efficient evidence-grounded scoring capability, PaSaMaster trains a dedicated lightweight Scorer model through knowledge distillation. The Scorer serves as the verification component of the Librarian: given a query-specific checklist and retrieved evidence chunks, it assigns checklist-level scores, generates evidence-grounded rationales, and produces a holistic relevance judgment for each candidate paper. To train this capability, we construct a corpus by first clustering

papers into multidisciplinary topic groups and then synthesizing natural-language search queries from each cluster. Then use the PaSaMaster retrieval system to produce noisy but deployment-matched candidate sets. A stronger teacher model then annotates each query–paper pair with checklist-level scores, evidence-grounded rationales, and holistic judgments. The resulting Scorer model allows Librarian agents to reproduce expert-style structured verification at much lower inference cost over large scientific corpora.

3. PaSaMaster-Bench

To evaluate whether literature retrieval systems can truly understand complex natural-language research intents, we introduce **PaSaMaster-Bench**, the first multidisciplinary benchmark designed for complex scientific literature search intents. Unlike conventional retrieval benchmarks that primarily evaluate keyword matching or short query–document relevance, PaSaMaster-Bench focuses on realistic research questions expressed as detailed natural-language intents. Each task requires a system to interpret a multi-constrained academic query, identify the underlying target paper set, and return a ranked list of real scientific papers that satisfy all specified conditions.

PaSaMaster-Bench contains **244 independent literature discovery tasks** across **38 scientific disciplines**. Each task is constructed around a complex search intent involving multiple explicit and implicit constraints, such as topical scope, methodological requirements, application scenarios, benchmark datasets, publication conditions, temporal restrictions, and exclusion criteria. The key design principle is that a paper is considered correct only if it satisfies the full intent expressed by the query, rather than merely matching isolated keywords or being broadly related to the topic. Therefore, strong performance on PaSaMaster-Bench directly indicates that a system can transform complex natural-language research needs into accurate target-paper retrieval.

3.1. Benchmark Construction

The construction of PaSaMaster-Bench follows a two-stage expert-driven pipeline. First, domain experts formulate complex natural-language literature search queries based on authentic research bottlenecks. For each query, experts also provide a constraint checklist that decomposes the intended search need into objective, verifiable criteria. These checklists define what it means for a paper to satisfy the user's intent.

Second, we build a comprehensive candidate pool for each query through omni-channel retrieval. Each query is searched across multiple strong systems, including web-enabled frontier LLMs(e.g., GPT-5.2, Gemini 3.1 Pro), PaSaMaster's native search engine, and traditional web search. Retrieved papers are first verified against the corpus, then deduplicated and organized into a unified candidate set. Domain experts then evaluate each candidate paper against the predefined checklist, assigning checkpoint-level judgments to verify whether the paper fully satisfies the query intent. Only papers that satisfy all required checkpoints are admitted into the ground-truth target set $\mathcal{P}^*$.

The benchmark is guided by four principles. First, intent fidelity: each task must reflect a realistic research intent rather than an artificial keyword query. Second, bounded recall: the target paper set must be sufficiently well-defined for expert annotation and objective evaluation. Third, authentic complexity: each query must require non-trivial interpretation of multiple constraints. Fourth, verifiable correctness: every ground-truth paper must be justified by checklist-based evidence. Together, these principles make PaSaMaster-Bench a direct test of whether retrieval systems can understand complex academic intents and retrieve the corresponding target literature.

3.2. Evaluation Protocol

Given a complex natural-language research query, a system is required to autonomously search, verify, and return a ranked list of papers:

$$\mathcal{P}_{\text{agent}} = [p_1, p_2, \ldots, p_k].$$

The returned list is compared against the expert-annotated target paper set $\mathcal{P}^*$. Because $\mathcal{P}^*$ is defined by strict checklist satisfaction, retrieval performance on PaSaMaster-Bench measures more than topical relevance: it measures whether the system correctly understood the user's full search intent and translated that understanding into the right set of papers.

We use standard retrieval metrics to evaluate this ability, all computed at a cutoff of $K = 20$:

- **Recall@K** measures whether the system can comprehensively recover the target papers implied by the query, i.e., the fraction of ground-truth papers that appear in the top-K returned results.
- **Precision@K** measures whether the top-K results satisfy the full expert-defined intent, i.e., the fraction of returned papers that are genuine ground-truth papers.
- **F1@K** summarizes the balance between comprehensively recovering target papers and avoiding papers that only partially match the intent, computed as the harmonic mean of Precision@K and Recall@K.
- **NDCG@K** further evaluates whether papers satisfying the intended criteria are ranked near the top, using a logarithmically discounted cumulative gain normalized against the ideal ranking.

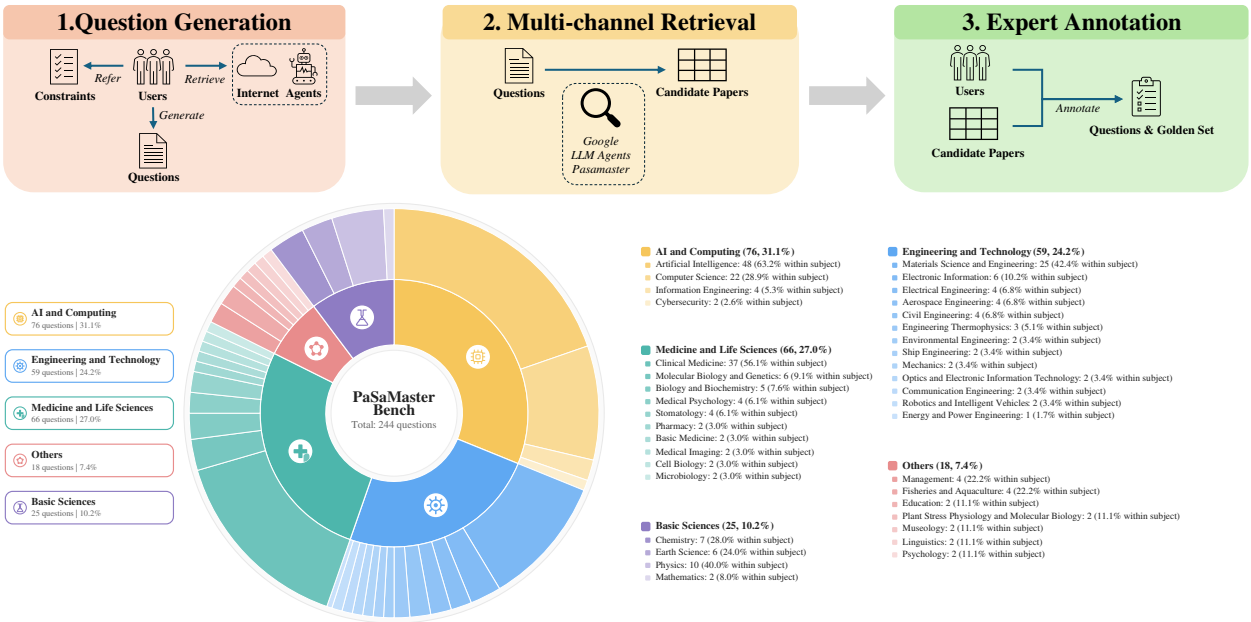

Figure 2. Overview of the PaSaMaster-Bench data curation pipeline, comprising three stages. (1) *Question Generation*: domain experts formulate complex natural-language search queries grounded in authentic research bottlenecks and supply a constraint checklist that decomposes the intent into objective, verifiable criteria. (2) *Multi-channel Retrieval*: each query is submitted to multiple strong retrieval systems — including web-enabled frontier LLMs, PaSaMaster's native search engine, and traditional web search — to build a comprehensive candidate pool. (3) *Expert Annotation*: domain experts evaluate each candidate paper against the predefined checklist and admit only those satisfying all required checkpoints into the ground-truth target set $\mathcal{P}^*$.

In addition to retrieval quality, we measure token usage and source hallucination rate. Token usage quantifies the cost of understanding and searching under complex constraints, while hallucination rate measures whether returned papers are real and verifiable. Together, these metrics evaluate the three central requirements of complex scientific literature discovery: *intent comprehension*, *source authenticity*, and *cost-efficient retrieval*.

4. Experiments and results

4.1. Experimental Setup

Baselines. We compare PaSaMaster with representative systems from the major paradigms of scientific literature retrieval. *Lexical retrieval systems* include Google Scholar (Google, n.d.), which represents keyword-centric search over indexed literature databases. *Semantic retrieval systems* include OpenScholar (Asai et al., 2024) and Bohrium Science Navigator (Zhang et al., 2025a), which improve over lexical matching by using semantic representations but still operate as passive query–document retrieval systems. *Generative LLMs* include DeepSeek-v3.2 (DeepSeek-AI et al., 2025), Kimi-K2.5 (Team et al., 2026), MiniMax-M2.7 (MiniMax et al., 2025), GLM-5 (Team et al., 2025), Gemini-3.1-pro (Google DeepMind, 2026), and GPT-5.2 (OpenAI, 2026). These models are equipped with Search and Visit tools and prompted to perform ReAct-style literature discovery before returning a final paper list. *Fixed-pipeline agentic retrieval systems* include Google Scholar Labs (Google, 2025), which grounds retrieval in verifiable tools but follows a predefined retrieve–read–answer workflow without iterative intent refinement from ranked evidence. PaSaMaster represents the *self-evolving agentic retrieval* paradigm, where ranked evidence is used to refine the search intent and guide subsequent retrieval rounds.

Evaluation Metrics. We evaluate each system by comparing its returned paper list with the expert-annotated ground-truth set in PaSaMaster-Bench. Retrieval quality is measured using Recall@20, Precision@20, F1-score@20, and NDCG@20. Recall@20 measures whether a system can recover the target papers implied by the complex search intent, while Precision@20 measures whether the returned top papers fully satisfy the expert-defined constraints. F1-score@20 summarizes the balance between completeness and correctness. NDCG@20 further evaluates ranking quality by assigning higher scores when ground-truth papers are placed closer to the top of the returned list.

In addition to retrieval quality, we report two practical reliability and efficiency metrics: hallucination rate and token usage cost. Hallucination rate measures the proportion of returned papers that cannot be verified as real scientific

Table 2. Performance comparison of PaSaMaster against other methods. With zero hallucinations guaranteed, PaSaMaster achieves strong recall and ranking (NDCG) with low cost.

Method	NDCG@20 (%)	Recall@20 (%)	Precision@20 (%)	F1-score@20 (%)	Hallucination	Cost ($)
Lexical Retrieval Systems						
Google Scholar	2.07	1.69	1.48	1.39	0	–
Semantic Retrieval Systems						
OpenScholar	14.61	11.68	8.52	7.92	0	–
Bohrium Science Navigator	22.39	19.37	12.50	12.26	0	–
Generative LLMs						
DeepSeek-v3.2	35.82	24.76	15.35	15.56	12.94	0.28
Kimi-K2.5	37.80	28.08	16.95	17.36	26.59	0.16
MiniMax-M2.7	30.70	24.23	14.42	15.11	32.66	0.18
GLM-5	35.89	28.99	16.93	18.18	21.64	0.56
Gemini-3.1-pro	31.34	21.30	11.68	12.48	27.54	0.38
GPT-5.2	31.59	25.32	16.82	16.69	5.65	6.06
Fixed-Pipeline Agentic Retrieval						
Google Scholar Labs	30.54	29.01	18.79	18.87	0	–
Self-Evolving Agentic Retrieval						
PaSaMaster	**39.52**	**33.24**	**23.46**	**23.00**	**0**	**0.05**

Table 3. **Performance comparison of different retrieval and generation systems.** Generative retrieval methods are prone to hallucinations, while database-backed retrieval systems return only indexed records and thus achieve zero hallucination.

Category	Method	Title	Author	Date	Link	All
Keywords-based Retrieval Systems	Google Scholar	0	0	0	0	0
Semantic Retrieval Systems	OpenScholar	0	0	0	0	0
	Bohrium Science Navigator	0	0	0	0	0
Generative LLMs	Gemini-3.1	2.52	17.79	9.26	4.14	27.54
	DeepSeek-v3.2	1.22	4.09	8.12	2.56	12.94
	GLM-5	2.97	7.67	13.12	5.80	21.64
	GPT-5.2	0.46	0.99	3.03	1.95	5.65
	MiniMax-M2.7	7.82	11.24	21.80	9.87	32.66
	Kimi-K2.5	4.06	12.90	15.17	5.85	26.59
Fixed-Pipeline Agentic Retrieval	Google Scholar Labs	0	0	0	0	0
Self-Evolving Agentic Retrieval	**PaSaMaster**	0	0	0	0	0

sources. Cost is computed from the total input and output tokens consumed during inference, priced according to the corresponding model rates. Note: The cost for Gemini 3.1 is estimated using a 20-question sample.

4.2. Results

Table 2 reports the main results on PaSaMaster-Bench. Overall, PaSaMaster achieves the best retrieval quality while maintaining zero source hallucination and low computational cost. These results support the three central claims of our system: self-evolving retrieval improves understanding of complex natural-language search intents, intent–paper relevance ranking prevents hallucinated sources, and planning–retrieval separation enables cost-efficient large-scale literature discovery.

Self-Evolving Retrieval Improves Complex Intent Understanding. PaSaMaster achieves the highest retrieval performance across all main quality metrics, with an NDCG of 39.52, Recall of 33.24, Precision of 23.46, and F1-score of 23.00. This demonstrates that PaSaMaster is better able to recover the target papers implied by complex multi-constraint research intents. Compared with Google Scholar, PaSaMaster improves F1-score from 1.39 to 23.00, a 16.5× improvement, showing the severe limitation of keyword-centric retrieval under complex natural-language queries. Compared with semantic retrieval systems, PaSaMaster also substantially outperforms OpenScholar and Bohrium Science Navigator, indicating that passive semantic matching is still insufficient for queries requiring constraint reasoning and intent refinement. Among generative LLMs, the strongest F1-score baseline is GLM-5 with 18.18, while the fixed-pipeline agentic retrieval baseline Google Scholar Labs achieves 18.87. PaSaMaster reaches 23.00, improving

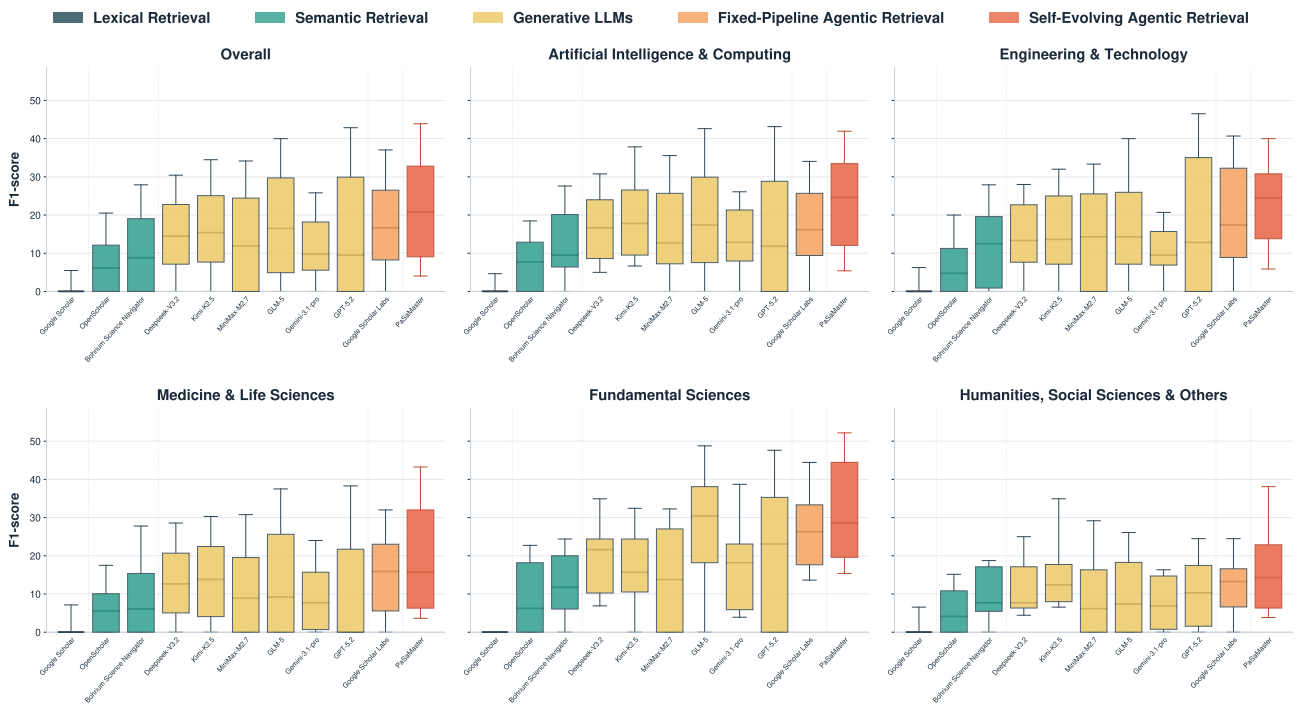

Figure 3. Per-discipline F1-score comparison across all methods. PaSaMaster consistently achieves the highest or comparable F1-scores across all subject domains.

over these strongest baselines by 26.5% and 21.9%, respectively. This advantage suggests that using ranked evidence to identify coverage gaps, refine intents, and guide follow-up searches provides a stronger mechanism for complex literature discovery than either generative answer synthesis or fixed-pipeline agentic retrieval.

Intent–Paper Ranking Eliminates Source Hallucination. PaSaMaster achieves 0% hallucination while maintaining the strongest retrieval quality. This result directly supports our design choice of treating literature discovery as intent–paper relevance ranking rather than generation. In contrast, generative LLM baselines exhibit substantial hallucination rates, including 32.66% for MiniMax-M2.7, 27.54% for Gemini-3.1, 26.59% for Kimi-K2.5, 21.64% for GLM-5, 12.94% for DeepSeek-v3.2, and 5.65% for GPT-5.2. These results show that even frontier LLMs with search and visit tools remain vulnerable to fabricating or misreporting scientific sources. Table 3 further shows that hallucinations arise from multiple citation fields, including title, author, date, and link errors. By contrast, PaSaMaster ranks only papers retrieved from verified corpora and grounds relevance judgments in original paper evidence, thereby ensuring zero hallucination in source information.

Planning–Retrieval Separation Reduces Cost Without Sacrificing Quality. PaSaMaster achieves this performance at a cost of only \$0.05 per query. This is far below GPT-5.2 at \$6.06, GLM-5 at \$0.56, Gemini-3.1-pro at \$0.38, and DeepSeek-v3.2 at \$0.28. In particular, PaSaMaster outperforms GPT-5.2 in F1-score by 37.8% while using only about 1% of its computational cost. This confirms the benefit of separating high-level planning from large-scale retrievalenabling PaSaMaster to maintain high-quality retrieval at substantially lower cost.

Consistent Gains Across Disciplines. Figure 3 reports the per-discipline F1-score comparison across the 38 scientific disciplines in PaSaMaster-Bench. PaSaMaster consistently achieves strong performance across diverse areas, including Artificial Intelligence and Computing, Engineering and Technology, Medicine and Life Sciences, Basic Sciences, and interdisciplinary fields. This cross-domain consistency indicates that the gains do not come from a narrow domain-specific advantage. Instead, they reflect the generality of the three design principles: self-evolving retrieval supports complex intent understanding, evidence-grounded ranking ensures source authenticity, and planning–retrieval separation enables scalable retrieval across heterogeneous scientific domains.

5. Conclusion

We introduced **PaSaMaster**, a self-evolving agentic literature retrieval system for complex natural-language scientific search intents. PaSaMaster addresses the key tension in modern literature discovery: understanding rich academic

search intents while ensuring that every returned source is real and verifiable. Its core contribution is a *self-evolving* literature retrieval paradigm that moves beyond one-shot query–document matching and generative citation synthesis. PaSaMaster iteratively analyzes search intent, uses ranked evidence to identify coverage gaps, refines the retrieval direction, and guides follow-up searches to better align with complex research needs. By ranking verified papers rather than generating citations, it eliminates source hallucination, while its separation of frontier-LLM planning from large-scale retrieval and lightweight relevance scoring enables high-quality retrieval at substantially lower cost. Experiments on PaSaMaster-Bench across 38 disciplines show that PaSaMaster provides researchers with more accurate, trustworthy, and cost-efficient literature discovery.

Funding Disclosure

This work was supported by the Shanghai Jiao Tong University–DP Technology Super Scientific Research Platform Strategic Cooperation Agreement (AI-X Research Institute).

Acknowledgements

The authors thank the developers of the Science Navigator team at DP Technology for their useful tools and prior efforts in scientific literature retrieval.

References

Ajith, A., Xia, M., Chevalier, A., Goyal, T., Chen, D., and Gao, T. Litsearch: A retrieval benchmark for scientific literature search. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 15068–15083. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.emnlp-main.840.

Asai, A., He, J., Shao, R., Shi, W., Singh, A., Chang, J. C., Lo, K., Soldaini, L., Feldman, S., D'arcy, M., Wadden, D., Latzke, M., Tian, M., Ji, P., Liu, S., Tong, H., Wu, B., Xiong, Y., Zettlemoyer, L., Neubig, G., Weld, D., Downey, D., tau Yih, W., Koh, P. W., and Hajishirzi, H. Openscholar: Synthesizing scientific literature with retrieval-augmented lms, 2024. URL http://arxiv.org/abs/2411.14199.

Bornmann, L. and Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11):2215–2222, 2015. doi: 10.1002/asi.23329.

DeepSeek-AI, Liu, A., Mei, A., Lin, B., Xue, B., Wang, B., Xu, B., Wu, B., Zhang, B., Lin, C., Dong, C., Lu, C., Zhao, C., Deng, C., Xu, C., Ruan, C., Dai, D., Guo, D., Yang, D., Chen, D., Li, E., Zhou, F., Lin, F., Dai, F., Hao, G., Chen, G., Li, G., Zhang, H., Xu, H., Li, H., Liang, H., Wei, H., Zhang, H., Luo, H., Ji, H., Ding, H., Tang, H., Cao, H., Gao, H., Qu, H., Zeng, H., Huang, J., Li, J., Xu, J., Hu, J., Chen, J., Xiang, J., Yuan, J., Cheng, J., Zhu, J., Ran, J., Jiang, J., Qiu, J., Li, J., Song, J., Dong, K., Gao, K., Guan, K., Huang, K., Zhou, K., Huang, K., Yu, K., Wang, L., Zhang, L., Wang, L., Zhao, L., Yin, L., Guo, L., Luo, L., Ma, L., Wang, L., Zhang, L., Zhang, M., Zhang, M., Tang, M., Zhou, M., Huang, P., Cong, P., Wang, P., Wang, Q., Zhu, Q., Li, Q., Chen, Q., Du, Q., Xu, R., Ge, R., Zhang, R., Pan, R., Wang, R., Yin, R., Xu, R., Shen, R., Zhang, R., Lu, S., Zhou, S., Chen, S., Cai, S., Chen, S., Hu, S., Liu, S., Hu, S., Ma, S., Wang, S., Yu, S., Zhou, S., Pan, S., Zhou, S., Ni, T., Yun, T., Pei, T., Ye, T., Yue, T., Zeng, W., Liu, W., Liang, W., Gao, W., Zhang, W., Bi, X., Liu, X., Wang, X., Chen, X., Zhang, X., Nie, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yu, X., Yang, X., Li, X., Su, X., Lin, X., Zhang, Y., Xu, Y., Li, Y., Zhao, Y., Sun, Y., Wang, Y., Yu, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Wu, Y., Ou, Y., Wang, Y., Gong, Y., Zou, Y., Xiong, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Xie, Z., Zhang, Z., Hao, Z., Gou, Z., Ma, Z., Yan, Z., Shao, Z., Wu, Z., Li, Z., Gu, Z., Zhu, Z., Li, Z., Xie, Z., Gao, Z., and Pan, Z. Deepseek-v3.2: Pushing the frontier of open large language models, 2025. URL https://arxiv.org/abs/2512.02556.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., and Barabási, A.-L. Science of science. *Science*, 359 (6379):eaao0185, 2018. doi: 10.1126/science.aao0185.

Google. Scholar labs: An AI powered scholar search. Google Scholar Blog, 2025. URL https://scholar.googleblog.com/2025/11/scholar-labs-ai-powered-scholar-search.html.

Google. Google Scholar. https://scholar.google.com/, n.d. Accessed: 2026-05-08.

Google DeepMind. Gemini 3.1 Pro: A smarter model for your most complex tasks, 2026. URL https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/.

Gusenbauer, M. and Haddaway, N. R. What every researcher should know about searching—clarified concepts, search advice, and an agenda to improve finding in academia. *Research Synthesis Methods*, 12(2):136–147, 2021. doi: 10.1002/jrsm.1457.

He, Y., Huang, G., Feng, P., Lin, Y., Zhang, Y., Li, H., and E, W. Pasa: An llm agent for comprehensive academic paper search, 2025. URL https://arxiv.org/abs/2501.10120.

MiniMax, Chen, A., Li, A., Gong, B., Jiang, B., Fei, B., Yang, B., Shan, B., Yu, C., Wang, C., Zhu, C., Xiao, C., Du, C., Zhang, C., Qiao, C., Zhang, C., Du, C., Guo, C., Chen, D., Ding, D., Sun, D., Li, D., Jiao, E., Zhou, H., Zhang, H., Ding, H., Sun, H., Feng, H., Cai, H., Zhu, H., Sun, J., Zhuang, J., Cai, J., Song, J., Zhu, J., Li, J., Tian, J., Liu, J., Xu, J., Yan, J., Liu, J., He, J., Feng, K., Yang, K., Xiao, K., Han, L., Wang, L., Yu, L., Feng, L., Li, L., Zheng, L., Du, L., Yang, L., Zeng, L., Yu, M., Tao, M., Chi, M., Zhang, M., Lin, M., Hu, N., Di, N., Gao, P., Li, P., Zhao, P., Ren, Q., Xu, Q., Li, Q., Wang, Q., Tian, R., Leng, R., Chen, S., Chen, S., Shi, S., Weng, S., Guan, S., Yu, S., Li, S., Zhu, S., Li, T., Cai, T., Liang, T., Cheng, W., Kong, W., Li, W., Chen, X., Song, X., Luo, X., Su, X., Li, X., Han, X., Hou, X., Lu, X., Zou, X., Shen, X., Gong, Y., Ma, Y., Wang, Y., Shi, Y., Zhong, Y., Duan, Y., Fu, Y., Hu, Y., Gao, Y., Fan, Y., Yang, Y., Li, Y., Hu, Y., Huang, Y., Li, Y., Xu, Y., Mao, Y., Shi, Y., Wenren, Y., Li, Z., Li, Z., Tian, Z., Zhu, Z., Fan, Z., Wu, Z., Xu, Z., Yu, Z., Lyu, Z., Jiang, Z., Gao, Z., Wu, Z., Song, Z., and Sun, Z. Minimax-m1: Scaling test-time compute efficiently with lightning attention, 2025. URL https://arxiv.org/abs/2506.13585.

National Center for Biotechnology Information. Pubmed, 1996. URL https://pubmed.ncbi.nlm.nih.gov/.

OpenAI. Introducing GPT-5.4, 2026. URL https://openai.com/index/introducing-gpt-5-4/.

Team, G., Zeng, A., Lv, X., Zheng, Q., Hou, Z., Chen, B., Xie, C., Wang, C., Yin, D., Zeng, H., Zhang, J., Wang, K., Zhong, L., Liu, M., Lu, R., Cao, S., Zhang, X., Huang, X., Wei, Y., Cheng, Y., An, Y., Niu, Y., Wen, Y., Bai, Y., Du, Z., Wang, Z., and Zhu, Z. Glm-4.5: Agentic, reasoning, and coding (ARC) foundation models, 2025. URL https://arxiv.org/abs/2508.06471.

Team, K., Bai, Y., Bao, Y., Charles, Y., Chen, C., Chen, G., Chen, H., Chen, H., Chen, J., Chen, N., Chen, R., Chen, Y., Chen, Y., Chen, Y., Chen, Z., Cui, J., Ding, H., Dong, M., Du, A., Du, C., Du, D., Du, Y., Fan, Y., Feng, Y., Fu, K., Gao, B., Gao, C., Gao, H., Gao, P., Gao, T., Ge, Y., Geng, S., Gu, Q., Gu, X., Guan, L., Guo, H., Guo, J., Hao, X., He, T., He, W., He, W., He, Y., Hong, C., Hu, H., Hu, Y., Hu, Z., Huang, W., Huang, Z., Huang, Z., Jiang, T., Jiang, Z., Jin, X., Kang, Y., Lai, G., Li, C., Li, F., Li, H., Li, M., Li, W., Li, Y., Li, Y., Li, Y., Li, Z., Li, Z., Lin, H., Lin, X., Lin, Z., Liu, C., Liu, C., Liu, H., Liu, J., Liu, J., Liu, L., Liu, S., Liu, T., Liu, W., Liu, Y., Liu, Y., Liu, Y., Liu, Y., Liu, Z., Lu, E., Lu, H., Lu, L., Luo, Y., Ma, S., Ma, X., Ma, Y., Mao, S., Mei, J., Men, X., Miao, Y., Pan, S., Peng, Y., Qin, R., Qin, Z., Qu, B., Shang, Z., Shi, L., Shi, S., Song, F., Su, J., Su, Z., Sui, L., Sun, X., Sung, F., Tai, Y., Tang, H., Tao, J., Teng, Q., Tian, C., Wang, C., Wang, D., Wang, F., Wang, H., Wang, H., Wang, J., Wang, J., Wang, J., Wang, S., Wang, S., Wang, S., Wang, X., Wang, Y., Wang, Y., Wang, Y., Wang, Y., Wang, Y., Wang, Z., Wang, Z., Wang, Z., Wei, C., Wei, Q., Wu, H., Wu, W., Wu, X., Wu, Y., Xiao, C., Xie, J., Xie, X., Xiong, W., Xu, B., Xu, J., Xu, L., Xu, S., Xu, W., Xu, X., Xu, Y., Xu, Z., Xu, J., Yan, J., Yan, Y., Yang, H., Yang, X., Yang, Y., Yang, Y., Yang, Z., Yang, Z., Yang, Z., Yao, H., Yao, X., Ye, W., Ye, Z., Yin, B., Yu, L., Yuan, E., Yuan, H., Yuan, M., Yuan, S., Zhan, H., Zhang, D., Zhang, H., Zhang, W., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, Z., Zhao, H., Zhao, Y., Zhao, Z., Zheng, H., Zheng, S., Zhong, L., Zhou, J., Zhou, X., Zhou, Z., Zhu, J., Zhu, Z., Zhuang, W., and Zu, X. Kimi k2: Open agentic intelligence, 2026. URL https://arxiv.org/abs/2507.20534.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T.-Y., Manrai, A., Marks, D., Ramsundar, B., Song, L., Sun, J., Tang, J., Veličković, P., Welling, M., Zhang, L., Coley, C. W., Bengio, Y., and Zitnik, M. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023. doi: 10.1038/s41586-023-06221-2.

White, R. W. and Roth, R. A. *Exploratory Search: Beyond the Query-Response Paradigm*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009. doi: 10.2200/S00174ED1V01Y200901ICR003.

Zhang, L., Chen, S., Cai, Y., Chai, J., Chang, J., Chen, K., Chen, Z. X., Ding, Z., Du, Y., Gao, Y., Gao, Y., Gao, J., Gao, Z., Gu, Q., Hong, Y., Huang, Y., Fang, X., Ji, X., Ke, G., Lei, Z., Li, X., Li, Y., Liao, R., Lin, H., Lin, X., Liu, Y., Liu, X., Liu, Z., Lu, J., Miao, T., Que, H., Sun, W., Wang, Y., Wu, B., Xue, T., Ye, R., Zeng, J., Zhang, D., Zhang, J., Zhang, L., Zhang, T., Zhang, W., Zhang, Y., Zhang, Z., Zheng, H., Zhou, H., Zhu, T., Zhu, X., Zhou, Q., and E, W. Bohrium + scimaster: Building the infrastructure and ecosystem for agentic science at scale, 2025a. URL https://arxiv.org/abs/2512.20469.

Zhang, Y., Khan, S. A., Mahmud, A., Yang, H., Lavin, A., Levin, M., Frey, J. G., Dunnmon, J., Evans, J., Bundy, A., Džeroski, S., Tegnér, J., and Zenil, H. Exploring the role

of large language models in the scientific method: from hypothesis to discovery. *npj Artificial Intelligence*, 1(1): 14, 2025b. doi: 10.1038/s44387-025-00019-5.

A. PaSaMaster Topic Taxonomy

The data construction pipeline uses a broad academic taxonomy with 19 top-level disciplines and 97 fine-grained research topics. On average, there are 5.1 fine-grained topics per discipline. The detailed discipline coverage is presented in Table 4.

Table 4. PaSaMaster Topic Taxonomy: Discipline Coverage

Discipline	Topics	Fine-grained topics
Geoscience	10	Geology; Geophysics; Geochemistry; Hydrology and Water Resources; Ocean Science; Atmospheric Science and Meteorology; Remote Sensing and GIS; Seismology and Disaster Geoscience; Geomorphology and Quaternary Geology; Resource and Energy Geology
Physics and Astrophysics	10	High Energy and Particle Physics; Nuclear Physics; Condensed Matter Physics; Atomic, Molecular, and Optical Physics; Statistical Physics and Complex Systems; Plasma Physics; Gravity and Cosmology; Astrophysics and Galaxy Evolution; Quantum Information and Quantum Computing; Soft Matter and Biophysics
Mathematics	10	Algebra; Geometry; Topology; Number Theory; Analysis; Partial Differential Equations; Probability and Stochastic Processes; Statistics; Optimization and Operations Research; Numerical Analysis and Scientific Computing
Agroforestry Science	5	Crop & Horticultural Science; Plant Protection & Pathology; Soil, Nutrients & Agro-ecosystems; Animal Science & Aquaculture; Smart Ag Engineering & Food Processing
Material Science	5	Structural & Composite Materials; Polymer & Soft Materials; Electronic, Semiconductor & Nano/Spin Materials; Energy & Battery Materials; Characterization, Failure & Reliability
Computer Science	5	AI & Machine Learning; Data Systems & Data Mining; Systems & Distributed Computing; Networks, Security & Privacy; Software Engineering & Reliability
Environmental Science and Ecology	5	Ecosystems & Biodiversity; Climate Change & Carbon Cycle; Pollution Chemistry & Remediation; Environmental Engineering & Resource Recovery; Monitoring, Modeling & Governance
Artificial Intelligence	14	LLM Agents & Tool-Using Systems; Multimodal AI and Perception-Language-Action; Computer Vision for Intelligent Agents; Natural Language Understanding and Reasoning; World Models & Model-Based Intelligence; General Intelligent Agents; Multi-Agent Systems; Embodied AI & Robot Learning; Autonomous Driving & Intelligent Transportation; Knowledge Representation & Neuro-Symbolic AI; Causal AI; Human-AI Interaction & Alignment; AI Safety, Robustness & Governance; Automated Scientific Discovery
Chemistry	5	Organic & Polymer Chemistry; Inorganic & Materials Chemistry; Physical & Computational Chemistry; Analytical Chemistry; Catalysis & Electrochemistry
Engineering Technology	5	Mechanical Engineering & Intelligent Manufacturing; Electrical/Electronics & Power Systems; Control, Robotics & Intelligent Transportation; Civil Engineering & Resilient Infrastructure; Chemical, Materials & Energy Engineering
Biology	5	Molecular & Cellular Biology; Genetics/Genomics & Bioinformatics; Microbiology & Immunology; Development, Evolution & Ecology; Neuroscience & Systems Biology
Medicine	5	Clinical Medicine & Surgery; Oncology & Precision Therapy; Cardiovascular & Vascular Medicine; Infectious Disease & Immunology; Medical Imaging/AI & Drug Development
Comprehensive Periodical	5	Interdisciplinary Research & Applications; Scientometrics & Research Evaluation; Science & Technology Policy; Data, Benchmarks & Research Infrastructure; Open Science & Methodology
Law	5	Public Law; Private Law; Criminal & Procedural Law; International & Economic/Financial Law; Technology, Data, IP & Environment
Psychology	5	Cognitive & Neuropsychology; Developmental & Educational Psychology; Social, Personality & Organizational Psychology; Clinical, Counseling & Health Psychology; Psychometrics & Methodology
Pedagogy	5	Curriculum & Instruction; Learning Sciences & Educational Technology; Educational Assessment & Measurement; Teacher Education & Professional Development; Education Policy & Management
Economics	5	Macroeconomics & Policy; Microeconomics, IO & Competition; Econometrics & Causal Methods; Finance & Financial Economics; Development, Labor & Public/International Economics
Management	5	Strategy, Innovation & Entrepreneurship; Organizational Behavior & Human Resources; Marketing & Analytics; Operations, Supply Chain & Project Management; Digital/Financial Management & Risk Governance
Humanities	5	Philosophy, Ethics & Technology; History & Archaeology; Linguistics & Literary Studies; Art & Cultural Studies; Communication, Media & Religious Studies