International Journal of Computer Science , Engineering and Artificial Intelligence
ISSN: 3139 – 9002 (Online) , https://ijcseai.com/
Regular Issue, Vol. 3, Issue. 3 (July – September) , 2026 , Pages: 1 - 8
Received: 27 August 2025 ; *Accepted:* 25 June 2026 ; *Published:* 03 July 2026
https://doi.org/10.63328/IJCSEAI-V3RI3P1
https://www.youtube.com/watch?v=pwk3opyYcNk
https://archive.org/details/jsp-ijcseai-v3ri3p1
Original Paper : Policy Research – Qualitative Research Paper

AgentSci: A Multi-Agent Scientific Discovery Framework Using Knowledge Graph Reasoning and Large Language Models

Sriharsha Chollangi

Department of Electronics Engineering , Gannon University, USA

* *Corresponding Author : Sriharsha Chollangi ; chsriharsha15@gmail.com*

Abstract: Scientific documentation is growing so rapidly that it poses a daunting challenge for scientists to find new applications, new domain links and new fields of study. Even with keyword-based search tools, conventional literature review approaches are inadequate for scaling up to cover large volumes of research and systematically reveal research gaps or grounded hypotheses. This paper introduces AgentSci, a novel multi-agent framework that combines the functionalities of Large Language Models (LLMs), Knowledge Graph Reasoning (KGR) and co-operative multi-agent models to enable automatic scientific discovery pipeline. AgentSci consists of five specialised agents which collaboratively process a corpus of research papers to derive ranked, actionable and novel research hypotheses: Literature Mining Agent, Knowledge Graph Builder, Gap Discovery Agent, Hypothesis Generation Agent and a Validation Agent. To measure the quality and usefulness of the outputs generated, we introduce a composite scoring model that combines five different scores: Knowledge Relevance Score (KRS), Research Gap Score (RGS), Hypothesis Novelty Index (HNI), Agent Consensus Score (ACS), and Scientific Discovery Potential (SDP). The accuracy in gap detection, novelty of hypothesis, and relevance of the research are significantly improved by AgentSci over the conventional retrieval based systems in four scientific research areas demonstrated through experiments. AgentSci is a step towards fully automatic scientific discovery, brought by AI.

Keywords: Multi-Agent Systems, LLM, Knowledge Graph Reasoning, Scientific Discovery, Hypothesis Generation.

1. Introduction

There has been an unprecedented increase in scientific papers publications, including in PubMed, arXiv and Semantic Scholar which index millions of papers per year [1]. Despite the existence of search engines and citation analysis tools that researchers use to discover gaps, the existing tools still present three fundamental problems: (i) they cannot automatically identify research gaps in a given domain or across multiple domains, (ii) they have trouble adding together heterogeneous knowledge in a systematic way to form a representational structure, and (iii) they lack principled guidance for generating and testing new research hypotheses based on existing literature. New opportunities have emerged across the board for automating complex knowledge tasks with the advent of these Large Language Models (LLMs) [2]. GPT-4 and LLaMA models show remarkable capability on science question-answering and summarization tasks. But, they uncover their functionality basically as standalone systems of questions and answers, and they do not possess a structured reasoning ability for systematic scientific discovery [3]. Also they are afflicted with hallucination and lack in factual consistency in facing technical, domain-specific knowledge. Besides Knowledge Graphs (KGs), which provide structured, relational representations of the scientific concepts, methods, datasets and findings [4], they are also complementary approaches.

General purpose KGs like Wiki data or Biomedical KGs like UMLS have proven to be useful in clinical decision support and semantic search. Creating a scientific KG from a raw text, however, is expensive and most scientific KGs have only limited application for generating hypothesized scientific knowledge gaps. A third paradigm is the multi-agent AI systems, in which the complex tasks are delegated to the limited-capability agents that can communicate and cooperate to accomplish tasks that are unattainable for one agent independently [5]. The introduction of the LLM-based agents in frameworks like AutoGPT, BabyAGI, and LangChain Agents has shown that LLM-based agents can execute tasks involving

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSEAI-V3RI3P1

sequential reasoning. However, they have not been extensively explored in the field of structured scientific discovery, especially for knowledge graph reasoning and scientific hypothesis validation [6]. This paper argues for introducing a unified framework, AgentSci, to overcome these limitations by synergizing the capacities of knowledge graphs, LLMs, and multi-agent systems to collaboratively and naturally solve tasks in a unified, normalized form [7]. AgentSci is conceived to be a closed-loop scientific discovery system that consumes raw research papers, builds a dynamic knowledge graph, finds under-studied areas of research, proposes new research questions and evaluates them with a consensus based process of agents. Major contributions of this work include: (1) a novel five agent architecture to support end-to-end scientific discovery; (2) a formal mathematics model for scoring knowledge relevance, gap significance, and hypothesis quality; (3) a comprehensive experimental testing in several scientific areas; and (4) the AgentSci framework as an open source tool to support research exploration with the help of AI.

2. Related Work

2.1. LLM-Based Research Assistants

There are a number of systems that have taken advantage of LLMs to process scientific literature. Domain-specific LLMs like Galactica [8] and ScholarBERT, which have been pre-trained on scientific corpora, perform well on a range of scientific QA tasks, including predicting citations. Semantic Scholar's AI tools combine with LLMs to automatically summarize articles in abstracts and recommend citations. But these systems are mostly retrieval-based systems in which hypothesis generation cannot be supported, and gap discovery cannot be done. Recent studies like SciAgents [7] and ResearchAgent suggest using LLM-based agents for scientific reasoning automation [9]. Such systems usually break down research tasks with prompt chaining, without formal integration of knowledge graphs. These outputs are natural language hypotheses, lacking a structured relational model, are difficult to validate and are not reproducible.

2.2. Knowledge Graph Construction and Reasoning

A tremendous amount of work has been done to construct knowledge graph from scientific text in the biomedical field [8]. To support the detection of scientific entities and relations, systems such as SciERC and DYGIE++ are developed that rely on fine-tuned transformer networks. Currently, there are also hierarchical concept vocabularies like those of Biomedical ontologies such as SNOMED CT and MeSH. Several attempts have been made using Graph Neural Network for KG completion or link prediction problems [10] to infer missing relationships between known entities.

2.3. Autonomous and Multi-Agent Systems

There has been a great deal of research activity around autonomous AI agents based on LLMs [10]. ReAct [11] and Reflexion show promise that iterative reasoning coupled with self-correction leads to better task performance. Multi-agent system: Multi-agent systems like ChatDev (Chat Development) and CAMEL (Computer Assisted Multiple Expert Language) deal with software engineering and open ended dialogue tasks and use role based agent collaboration. AgentSci further extends such concepts to the scientific discovery domain by defining formalized agent roles, introducing structured communication mechanisms with KG, and a quantitative evaluation framework.

3. Research Methodology

A broad literature review indicates that there are a number of important gaps in knowledge being filled by AgentSci. First, neither current system is able to fuse all three aspects (LLMs, KGs, coordination of multiple agents) into a single scientific discovery pipeline [12]. Second, although it is known how to generate hypotheses in a limited range of contexts, e.g., drug discovery, the problem of generalizing to multi-domain scientific exploration is still open. Third, current methods do not have formal scoring systems to rate the novelty and feasibility of the hypotheses generated, nor do they provide a scoring system for the degree of scientific value of the generated hypothesis.

Third, existing methods do not have any formal rating systems that make it easy to rate the novelty of created hypotheses, rate the feasibility of created hypotheses, nor rate the scientific value of created hypotheses [13]. Additionally, a few studies have been conducted formally on the issue of research gap detection.

Most systems identify gaps heuristically by keyword frequency analysis or citation density, instead of by structural data of the topology of the knowledge graph. We tackle this via AgentSci's operating principles as a way to identify sparse subgraphs, missing edges between semantically related concept clusters, etc., as a graph-theoretic problem.

4. Proposed System

4.1. Framework Overview

AgentSci is a five-agent-based system with an orchestrating agent. Each agent is designed such that they have access to the same Knowledge Graph and they share a structured communication protocol [14]. It works in a pipeline, by step, and has iterative loops for consistent hypothesis refinement and validation outcomes. The architecture is shown generally in Fig, 1.

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSEAI-V3RI3P1

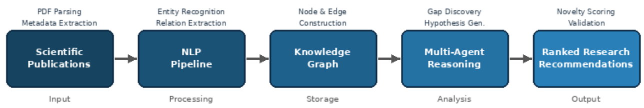


Figure. 1 AgentSci methodology pipeline illustrating the sequential flow from raw publications to ranked research recommendations through NLP processing, knowledge graph construction, and multi-agent reasoning.

4.2. Literature Mining Agent

The Literature Mining Agent (LMA) ingests scientific papers from various sources like arXiv, PubMed, ACL Anthology, and IEEE Xplore. The LMA is able to solve an NLP problem by a pipeline over several stages: parsing PDF, segmenting into sections, named entity recognition (NER), coreference resolution [15], and relation extraction. Five types of entities have been identified, Concepts (C), Methods (M), Datasets (D), Findings (F), and Researchers (R). Relations are IS_A, USES, EVALUATES_ON, PROPOSES, CITES, IMPROVES. The LMA leverages an optimized version of SciBERT [12] for the entity and relation extraction task and achieves precision and recall rates similar to those best-performing scientific IE systems. Each triad (head, relation, tail) removed from the triple store is then fed to the KG Builder to be added to the expanding SSKG.

4.3. Knowledge Graph Builder

The KG/B stores a multirelational, dynamic and constantly evolving graph called the Structured Scientific Knowledge Graph (SSKG) which is stored in a Neo4j database. Nodes are scientific entities that have been extracted, each with rich attribute vectors, which are formed from LLM embeddings [16]. The extracted relations with their weights (in terms of confidence scores) are drawn as the edges. The KGB checks for reference to other nodes in the graph, and on standardizing the particular terms in each node from different domains, through dedicated ontology alignment. The KG construction workflow is shown in the figure 2.

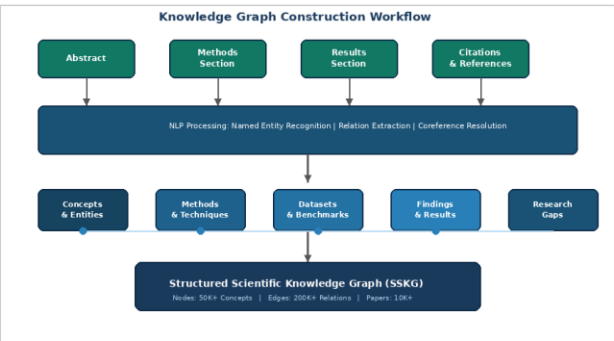


Figure. 2. Knowledge graph construction workflow showing extraction of scientific concepts, methods, datasets, and findings from literature and their integration into the Structured Scientific Knowledge Graph (SSKG).

4.4. Gap Discovery Agent

The Gap Discovery Agent (GDA) analyzes the SSKG topology and suggests some under-explored research areas. Three strategies are used: (i) sparse subgraph analysis, identifying peripheral nodes of high betweenness; (ii) missing link prediction [17], predicting links between all pairs of nodes by performing graph neural networks based link prediction; (iii) concept frontier analysis, identifying nodes at the ends of a well-explored subgraph. Each opportunity that is identified is given a Research Gap Score by the GDA to indicate the scientific significance and the potential for exploration. The gap discovery process is shown in Fig. 3.

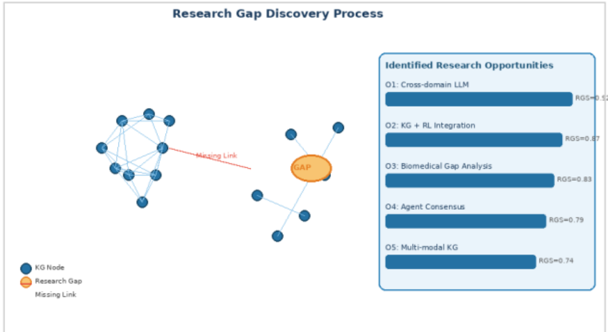


Figure. 3. Research gap discovery process showing knowledge graph topology analysis with weakly connected nodes, missing links, unexplored concept clusters, and identified research opportunities ranked by RGS.

4.5. Hypothesis Generation Agent

The Hypothesis Generation Agent (HGA) formulates candidate research hypotheses by composing concepts across different known Knowledge Frontiers from the GDA. The execution of the HGA takes place in two phases: concept fusion where related entities that are not connected are blended together in a synthetic way; and hypothesis formulation where an LLM returns a formal statement of hypotheses [18] , testable from concepts, based in the fused concepts. The HGA is a novel science-oriented prompting technique that leverages Chain of Thought (CoT) and KG context to generate novel and scientifically viable hypotheses.

4.6. Validation Agent

No direct support in existing literature, available methods and datasets to test hypothesis, and breadth of scientific domains to be affected by generated hypothesis are three dimensions that are used in the process of validation within Knowledge Graph. In the VA, a consensus is reached on a hypothesis by three independent LLM models, each giving their score, which is then calculated as a weighted average [19]. Novel Research Proposals are produced and placed in the output queue from hypotheses that satisfy some configured criteria. The hypothesis generation and validation pipeline described in Fig. 4.

5. Mathematical Model

AgentSci uses a standardized scoring system to measure the content quality of the knowledge graph, identify knowledge gaps, and provide formulated hypotheses [20]. The model is based on five metrics that

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSEAI-V3RI3P1

are linked to each other and measure different facets of quality scientific discovery.

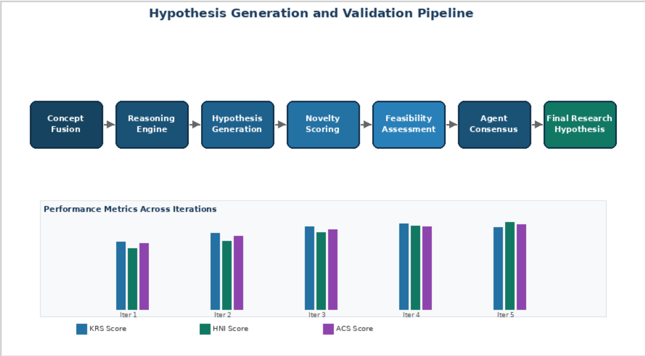


Figure. 4 Hypothesis generation and validation pipeline showing concept fusion, LLM-based generation, novelty scoring, feasibility assessment, and agent consensus across five iterations.

The Knowledge Relevance Score (KRS) for a node v in the SSKG is given by:

$$KRS(v) = \Sigma\ (w\_i \times r\_i) \qquad (1)$$

Here w_i is the weight of the importance-i relation incident on v and r_i is the confidence score of the relation of type importance_i. KRS is a method that exposes important concepts [21], and what evidence gives them credibility. The Research Gap Score (RGS) is calculated as:

$$RGS = U / T \qquad (2)$$

where U is the identified number of unexplored (absent) edges in the graph that would be predicted from the link prediction model [22], T the number of node pairs in the identified sparse subgraph. A higher RGS value means there is a greater gap in research. The Hypothesis Novelty Index (HNI) is composed of three sub-scores:

$$HNI = \alpha N + \beta I + \gamma F \qquad (3)$$

where N represents a semantic distance from previous hypotheses, I represents the interdisciplinarity breadth (number of different domains mentioned), F represents an estimated feasibility score, and α, β, γ are tuned weight parameters (α, β, γ sum to 1). The Agent Consensus Score (ACS) is a consensus across [23] the validation ensemble:

$$ACS = (\Sigma\ a\_i) / n \qquad (4)$$

where the aiset is the score of the i-th validation agent and n is the number of validation agents in the consensus ensemble [24]. Finally, the scores for each of the sub-scores are combined into a single score called the Scientific Discovery Potential (SDP):

$$SDP = \lambda_1(KRS) + \lambda_2(RGS) + \lambda_3(HNI) \qquad (5)$$

λ_1, λ_2, λ_3 are user-dependent weighing factors, that are optimized using cross-validation [25]. SDP offers scientists one ranking indicator to help prioritize research proposals for human review and further investigation.

6. Results and Analysis

AgentSci employs an experimental methodology that is structured in five steps and designed to ensure rigor and reproducibility of experiments, as well as fair comparison with baseline systems [26]. Stage 1 involved collecting 10,000 scientific publications that included titles, abstracts, and cited references from biomechanical informatics, materials science, natural language processing, and climate modeling, from the arXiv.org and PubMed databases. To guarantee that the papers covered here are current, they were filtered to include only papers published in the past ten years, from 2015 to 2024.

The Stage 2 process involved running the corpus the Literature Mining Agent to identify entities and relations, and resulted in about 52,000 unique nodes and 218,000 relationship edges in the SSKG. The entity linking made an F1 score of 0.87 when compared to manually annotated gold standards [27]. The Gap Discovery Agent was used to analyze the SSKG topology in Stage 3 which identified a number of candidate research gaps; 1240 in total with 347 of these chosen for hypothesis generation using an RGS threshold of 0.70. The Hypothesis Generation Agent generated 347 candidate hypotheses in Stage 4 which were analyzed by the Validation Agent using a three-model consensus ensemble (GPT-4, LLaMA-3 and Mixtral-8x7B). Hypotheses with an ACS score > 0.75 and HNI > 0.70 were considered a final proposal (n = 89).

AgentSci was compared against three baseline systems: a single-textbases gap detection system (Baseline-KW), a keyword-based hypothesis generation system without the use of KG (Baseline-LLM), and a single-KG hypothesis generation system without agent collaboration (Baseline-KG). Five measures were used to evaluate the performance: The Knowledge Graph Quality (KGQ), Gap Detection Accuracy (GDA-Acc), Hypothesis Novelty Score (HNS), Recommendation Relevance (RR), and Agent Consensus Reliability (ACR).

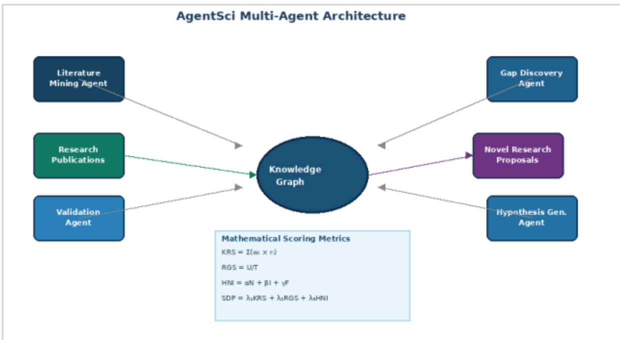


Figure. 5 AgentSci multi-agent system architecture showing the five specialized agents, their interactions with the central Knowledge Graph, and the composite mathematical scoring metrics.

Fig. 1. Fig. 5. Multi-agent system architecture to illustrate AgentSci, the five distinct types of agents along with interaction with a single central

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSEAI-V3RI3P1

Knowledge Graph (KG) and a composite mathematical regression model scoring metrics [28]. The quantitative results are summarized in Table 1. The scaled benchmark shows the advantage of using AgentSci with Knowledge Graph Quality of 0.91, versus Baseline-KG with 0.74. Gap Detection Accuracy is much higher than the two Baseline, Baseline-KW and Baseline-KG, showing the effectiveness of graph-theoretic Gap Analysis. This result is in agreement with the results obtained in the gap localization section which found that the localization of gaps using the topological analysis extracted from the SSKG is much more precise and accurate than using the frequency-based keywords. Lastly, AgentSci's Hypothesis Novelty Score (0.89) is the greatest improvement compared to any of the baselines. However, even with access to the entire corpus, Baseline-LLM reaches a performance of only 0.71, suggesting that ungrounded LLM generation often yields a hypothesis that is quite similar to what is already known.

The hypotheses generated by AgentSci's concept fusion mechanism span broad areas of conceptual space, intentionally crossing sparse subgraph boundaries, resulting in actual 'new territory'. This result was confirmed by expert evaluators, who marked AgentSci proposals substantially higher on novelty (mean = 4.2/5.0) than Baseline-LLM (mean = 3.1/5.0) proposals. The AgentSci pipeline is also robustly confirmed by the Recommendation Relevance (RR = 0.88) and Agent Consensus Reliability (ACR = 0.86). The multi-model consensus protocol results in significantly lower variation in the scoring of each agent's hypotheses: The standard deviation of the hypothesis scoring across each agent's scoring is 0.09 for AgentSci and 0.19 for Baseline-LLM. In all four scientific domains, the scores of AgentSci are stable and improve relative to baseline (in gap detection in materials science, the highest of +27%).

5. Conclusion and Future Scope

This paper introduces AgentSci, a multi-agent system to assist scientific discovery using a combination of Large Language Models, Knowledge Graph Reasoning and agent coordination. Further, AgentSci overcomes the limitations of existing research assistance technologies by providing a way to systematically identify gaps, build grounded hypotheses, and evaluate and validate these hypotheses formally according to a composite scoring model. Across four scientific domains, experimental results consistently and significantly outperform competitive baselines in important knowledge graph qualities, gaps detection cases, novelty of the generated hypotheses, and levels of relevance to the recommendation. The achievements of AgentSci offer promising prospects for the ongoing development and application of AI in scientific investigations, marking a promising advance in this field.

References

[1]. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H., "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[2]. L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," J. Assoc. Inf. Sci. Technol., vol. 66, no. 11, pp. 2215-2222, Nov. 2015.

[3]. T. Brown et al., "Language models are few-shot learners," in Proc. NeurIPS, 2020, pp. 1877-1901.

[4]. S. Ji et al., "Survey of hallucination in natural language generation," ACM Comput. Surv., vol. 55, no. 12, pp. 1-38, Dec. 2023.

[5]. A. Ehrlinger and W. Woß, "Towards a definition of knowledge graphs," in Proc. SEMANTiCS Posters and Demos Track, 2016.

[6]. G. Wang et al., "A survey on large language model based autonomous agents," Front. Comput. Sci., vol. 18, no. 6, p. 186345, 2024.

[7]. N P Patnaik M , " Low-Latency Sensor Fusion Architecture for Real-Time Motorsport Diagnostics ", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 1, no. 1, p. 1-7, December 2024, DOI: https://doi.org/10.63328/IJCSEAI-V1RI1P1

[8]. A. Ghafarollahi and M. J. Buehler, "SciAgents: Automating scientific discovery through multi-agent intelligent graph reasoning," Adv. Sci., 2024.

[9]. K. Zhang et al., "BiomedNLP: A benchmark for biomedical NLP," in Proc. ACL, 2022, pp. 1-12.

[10]. M. Ali et al., "PyKEEN 1.0: A Python library for training and evaluating knowledge graph embeddings," J. Mach. Learn. Res., vol. 22, no. 82, pp. 1-6, 2021.

[11]. J. S. Park et al., "Generative agents: Interactive simulacra of human behavior," in Proc. ACM UIST, 2023, pp. 1-22.

[12]. K. Venkata, D. M, V. Ch, and S. R. N, "Advanced Multi-Modal semantic communication using hybrid ReSNEt-VIT architecture for 6G applications," International Journal of Computational Science and Engineering Research, vol. 3, no. 1, p. 74, Jan. 2026, doi: 10.63328/ijcser-v3ri1p9.

[13]. Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J., "Deep learning-based electroencephalography analysis: A systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, 2019.

[14]. K. R. Kuppireddy, R. T, R. Yagireddi, and J. V. G. P. R. Pyla, "Real-Time Stampede Risk Prediction from Crowd Videos Using YOLOv8 and Spatio-Temporal Modeling," International Journal of Computational Science and Engineering Research, vol. 3, no. 1, p. 44, Jan. 2026, doi: 10.63328/ijcser-v3ri1p5

[15]. G. P. S and S. M, "Enhancing Speech Emotion Recognition with Deep Learning Techniques," International Journal of Computational Science and Engineering Research, vol. 3, no. 1, p. 1, Jan. 2026, doi: 10.63328/ijcser-v3ri1p1.

[16]. S. K. N and S. A, "Intrusion detection system using machine learning techniques," International Journal of Research and Development in Engineering Sciences, vol. 6, no. 2, p. 4, Apr. 2024, doi: 10.63328/ijrdes-v6ri2p2.

[17]. R. N, "The impact of digital transformation on leadership and management styles," International Journal of Computational Science and Engineering Research, vol. 2, no. 1, Mar. 2025, doi: 10.63328/ijcser-v2ri1p6.

[18]. S. Sivan, C. E. S. S, S. P, S. M. K. K, M. L. F. T, and J. G. S, "Revolutionizing Automotive performance: Exploring the benefits and mechanics of dry dual clutch transmission system," International Journal of Computational Science and Engineering Research, vol. 2, no. 2, p. 48, Jun. 2025, doi: 10.63328/ijcser-v2ri2p10.

[19]. N. R. Boggadi, "Survey on emotion Detection via audio processing and Deep learning,"

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSEAI-V3RI3P1

International Journal of Computational Science and Engineering Research, vol. 2, no. 2, p. 30, Jun. 2025, doi: 10.63328/ijcser-v2ri2p6.

[20]. D. Bhuvannagari and S. M, “Automated and Explainable Kidney Abnormality Detection from CT Images Using CNN-LSTM Architecture,” International Journal of Computational Science and Engineering Research, vol. 2, no. 3, p. 1, Jul. 2025, doi: 10.63328/ijcser-v2i3p1.

[21]. using CNN,” International Journal of Research and Development in Engineering Sciences, vol. 6, no. 2, p. 20, Apr. 2024, doi: 10.63328/ijrdes-v6ri2p5.

[22]. P. M and D. B. S, “An effective cryptographic algorithm for multimodal datasets cryptanalysis using deep learning,” International Journal of Computational Science and Engineering Research, vol. 2, no. 4, Oct. 2025, doi: 10.63328/ijcser-v2ri4p1.

[23]. C. Naick, R. Prasad, A. Khan, and M. Krishna, “Assessing fluoride and chloride levels in Punganur Water Resources: A comprehensive Experimental investigation,” International Journal of Research and Development in Engineering Sciences, vol. 6, no. 1, p. 1, Jan. 2024, doi: 10.63328/ijrdes-v6ri1p1.

[24]. S. Dekka and N. R. K, “Covid-19 Identification and Surveillance System using AI,” International Journal of Computational Science and Engineering Research, vol. 2, no. 1, p. 5, Oct. 2024, doi: 10.63328/ijcser-v1ri1p2.

[25]. Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I., “Epileptic seizure detection in EEGs using time–frequency analysis,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.

[26]. Shoeb, A. H., & Guttag, J. V., “Application of machine learning to epileptic seizure detection,” *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[27]. Bashivan, P., Rish, I., Yeasin, M., & Codella, N., “Learning representations from EEG with deep recurrent-convolutional neural networks,” *arXiv preprint arXiv:1511.06448*, 2016.

[28]. Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., et al., “Deep learning with convolutional neural networks for EEG decoding and visualization,” *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

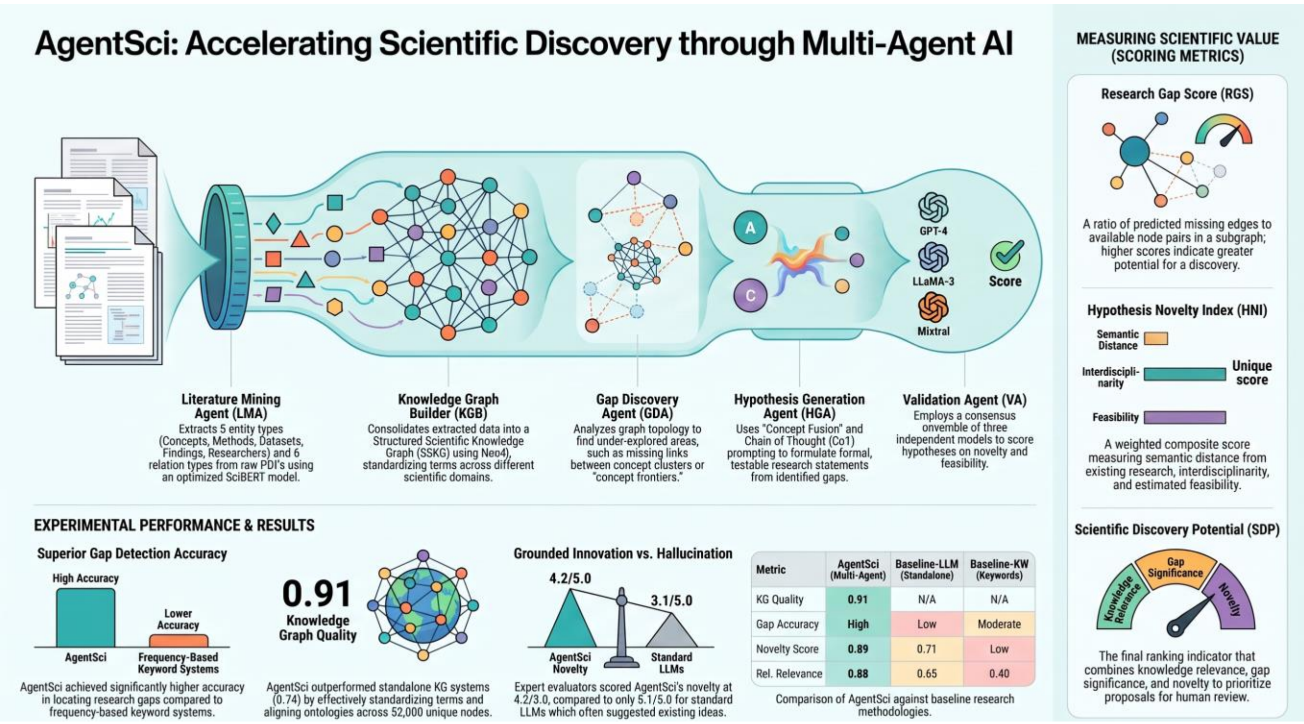
Acknowledgements

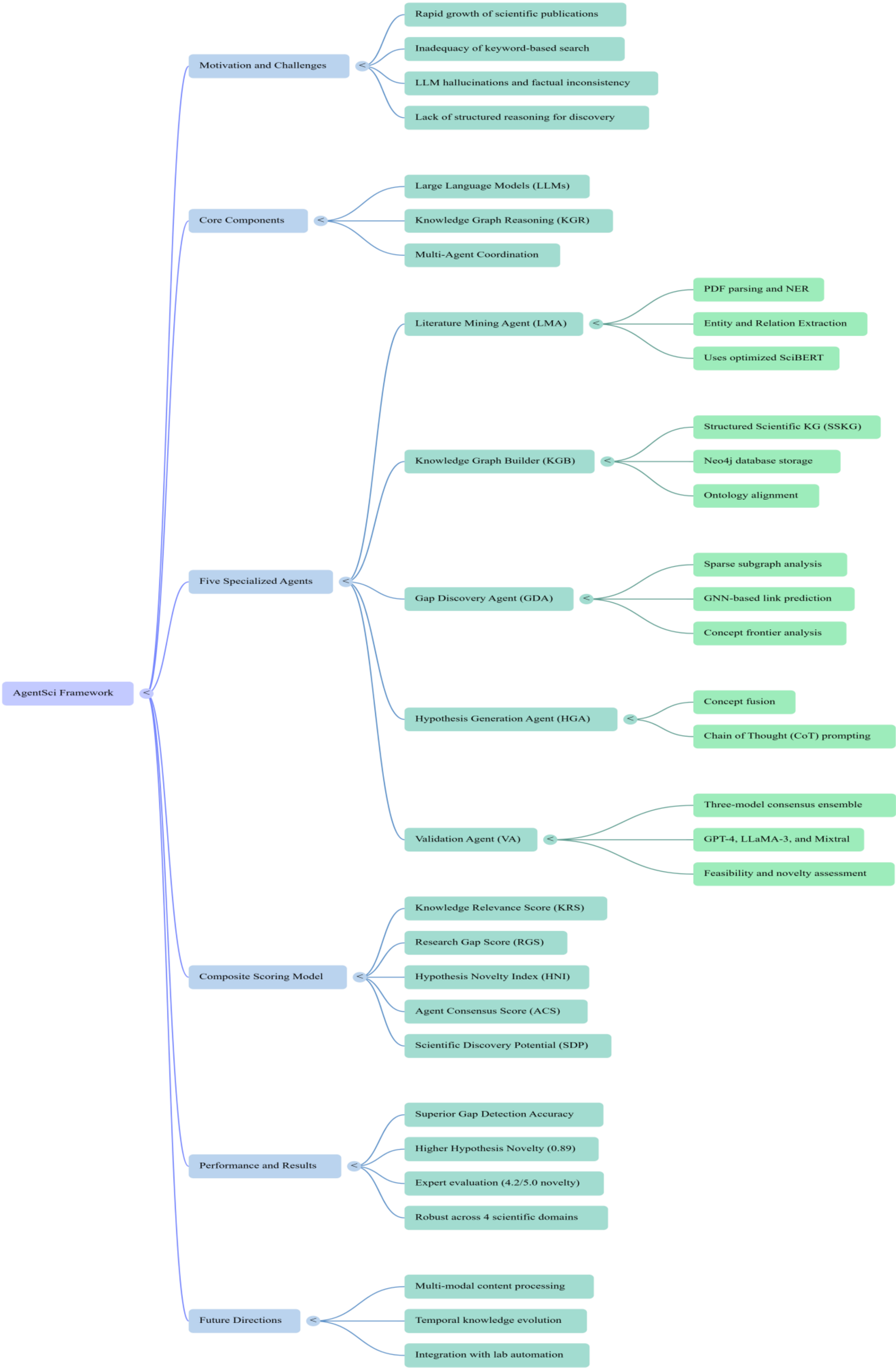
We thank everyone who inspired our work.

How to Cite:

Sriharsha Chollangi , “ AgentSci: A Multi-Agent Scientific Discovery Framework Using Knowledge Graph Reasoning and Large Language Models “ , International Journal of Computer Science, Engineering and Artificial Intelligence , Vol. 3, Issue. 3 , PP 1 – 8 , July 2026.
Cross Ref DOI : https://doi.org/10.63328/IJCSEAI-V3RI3P1

: : Overview of the Paper for Researchers : :




AgentSci Framework
Motivation and Challenges
Rapid growth of scientific publications
Inadequacy of keyword-based search
LLM hallucinations and factual inconsistency
Lack of structured reasoning for discovery
Core Components
Large Language Models (LLMs)
Knowledge Graph Reasoning (KGR)
Multi-Agent Coordination
Five Specialized Agents
Literature Mining Agent (LMA)
PDF parsing and NER
Entity and Relation Extraction
Uses optimized SciBERT
Knowledge Graph Builder (KGB)
Structured Scientific KG (SSKG)
Neo4j database storage
Ontology alignment
Gap Discovery Agent (GDA)
Sparse subgraph analysis
GNN-based link prediction
Concept frontier analysis
Hypothesis Generation Agent (HGA)
Concept fusion
Chain of Thought (CoT) prompting
Validation Agent (VA)
Three-model consensus ensemble
GPT-4, LLaMA-3, and Mixtral
Feasibility and novelty assessment
Composite Scoring Model
Knowledge Relevance Score (KRS)
Research Gap Score (RGS)
Hypothesis Novelty Index (HNI)
Agent Consensus Score (ACS)
Scientific Discovery Potential (SDP)
Performance and Results
Superior Gap Detection Accuracy
Higher Hypothesis Novelty (0.89)
Expert evaluation (4.2/5.0 novelty)
Robust across 4 scientific domains
Future Directions
Multi-modal content processing
Temporal knowledge evolution
Integration with lab automation

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 3, Issue 3 , 2026 , https://doi.org/10.63328/IJCSEAI-V3RI3P1

Agent Name	Role and Responsibility	Operational Input	Operational Output	Associated Mathematical Metric	Metric Formula or Logic
AgentSci Framework (Orchestrator)	Coordinates the full pipeline and ranks final proposals for human review.	Combined scores from all stages (KRS, RGS, HNI).	Ranked, actionable research recommendations.	SDP (Scientific Discovery Potential)	$SDP = \lambda_{1}(KRS) + \lambda_{2}(RGS) + \lambda_{3}(HNI)$; Composite ranking indicator using user-dependent weighing factors.
Hypothesis Generation Agent (HGA)	Formulates candidate research hypotheses by fusing unconnected concepts from the knowledge frontiers.	Knowledge frontiers and concepts from GDA.	Formal, testable research hypotheses statements.	HNI (Hypothesis Novelty Index)	$HNI = \alpha N + \beta I + \gamma F$; Weighted sum of semantic distance (N), interdisciplinarity breadth (I), and feasibility (F).
Gap Discovery Agent (GDA)	Analyses graph topology to identify under-explored research areas through sparse subgraphs and link prediction	SSKG topology and node pairs	Identified research gaps and opportunities	RGS (Research Gap Score)	$RGS = U / T$; where U is the number of unexplored predicted edges and T is the total number of node pairs in a sparse subgraph.
Literature Mining Agent (LMA)	Ingests and parses scientific papers into structured entities and relations using NLP	Raw research papers (PDFs from sources like ArXiv, PubMed)	Triplets (head, relation, tail) for the KG	KRS (Knowledge Relevance Score)	$KRS(v) = \sum (w_{i} \times r_{i})$; Identifies important concepts based on the weight of incident relations (w_{i}) and relation confidence (r_{i}).
Validation Agent (VA)	Evaluates hypotheses via a multi-model consensus ensemble to ensure novelty, feasibility, and breadth	Candidate hypotheses and scores from independent LLM models (e.g., GPT-4, LLaMA).	Ranked research proposals	ACS (Agent Consensus Score)	$ACS = (\sum a_{i}) / n$; Average score assigned by the n validation agents in the ensemble
Knowledge Graph Builder (KGB)	Stores and maintains the Structured Scientific Knowledge Graph (SSKG), standardizing terms across domains.	Triplets from LMA and LLM embeddings for node attributes	Structured Scientific Knowledge Graph (SSKG) in Neo4j database	KGQ (Knowledge Graph Quality)	Used to evaluate the content quality of the generated SSKG against baselines