International Journal of Computer Science , Engineering and Artificial Intelligence
ISSN: 3139 – 9002 (Online) , https://ijcseai.com/
Regular Issue, Vol. 1, Issue. 1 (October – December) , 2024 , Pages: 14 - 19
Received: 21 June 2024 ; *Accepted:* 28 September 2024 ; *Published:* 14 October 2024
https://doi.org/10.63328/IJCSEAI-V1RI1P3
https://www.youtube.com/watch?v=XPLQLt9S-ZM
https://archive.org/details/jsp-ijcseai-v1ri1p3
Original Paper : Applied Research – Experimental Research Paper

Collaborative Swarm Intelligence Framework for Multi-Agent Scientific Discovery Systems

Chekuri Nalini [1] , Yam Krishna Poudel [2]

[1] Department of Electronics and Communication Technology, Mohan babu University, Tirupathi, Andhra Pradesh, India; nalini.chekuri@mbu.asia

[2] Department of Electrical and Electronics Engineering, Nepal Engineering College, Changunarayan, Bhaktapur, Nepal; yamkp@nec.edu.np

* ***Corresponding Author***: C. Nalini ; *nalini.chekuri@mbu.asia*

Abstract: Being a collaborative, iterative process, scientific discovery requires synthesis of large amounts of literature, the development of new theories or hypotheses, the design of new experiments, and extensive validation. Most of the existing automated research systems are centralized or single-agent pipelines whose ability is only applicable for parallel exploration, knowledge sharing, and consensus formation is limited. In this paper, a Collaborative Swarm Intelligence Framework (CSIF) for multi-agent scientific discovery is proposed where the rapidity and diversity of scientific discovery is accelerated by coordinating and sharing the knowledge of a swarm of specialized and autonomous research agents containing agents for literature search, hypothesis generation, experimental design, data analysis and validation. By relying on stigmergy with coordination and weighted consensus, it allows emergent collective intelligence, while avoiding any centralized bottlenecks as constructed from a swarm architecture. Novelty detection rates achieved are 98%, accuracy of the hypothesis of 96%, consensus quality score of 97% and the hypothesis validation classifier AUC of 0.96, significantly exceeding scores for single-agent, RAG-based, and hierarchical multi-agent baselines on experimental evaluation on arXiv metadata, Semantic Scholar corpora, as well as common scientific benchmark datasets.

Keywords: Swarm Intelligence, Multi-Agent Systems, Consensus Mechanisms, Hypothesis Generation.

1. Introduction

Enormous acceleration in science knowledge creation is one of the most paramount 21st century challenges. The quantities of published scientific work has been increasing rapidly, doubling about once every 9 years, and currently more than 4 million of scientific articles are published per year, irrespective of their discipline [1]. Cognitive limits of individual researchers and even large teams make it virtually impossible to keep track of this evolving stream of literature, to determine which fields of research are unexplored, to develop ambitious theories and hypotheses based on an adequate understanding of the previous research, and to evaluate individual experiments and results against this vast body of knowledge. Automated systems used for scientific discovery complement and expand the reach of human scientists by efficiently exploring a large, hypothesis space in a systematic and fast manner.

Swarm intelligence [2] offers an interesting theoretical approach towards such automation. Swarm Intelligence systems are inspired by the collective problem-solving capabilities of the social insects including ants and bees' colonies and swarms, which are built from many relatively simple autonomous agents that communicate locally. Individual agents interact locally with their own rules of interaction, and indirectly share information through local environment signals, giving rise to global behaviors that are adaptive, fault-tolerant, and solve problems which are unsolvable for centralized architectures. Following this line of thought, the scientific hypothesis space can be explored more comprehensively and efficiently by a swarm of special purpose research agents that are able to exchange and inform both their own and others' hypotheses and arrive on consensus conclusions than would be a single agent or centralized system. Current automated research systems are deficient in a number of important ways. The sequential exploration capacity of a single computational

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 1, Issue 1 , 2024 , https://doi.org/10.63328/IJCSEAI-V1RI1P3

thread is the major limitation of single-agent systems, and they do not have mechanisms for diversified parallel hypothesis generation. While retrieval-augmented generation (RAG) systems work well for improving factual grounding, they are, however, constrained by having only one retrieval-generation cycle without cross-agent knowledge synthesis.

The hierarchical multi-agent system provides specialization of the agents while it also establishes an environment's communication structure that is fixed and difficult to adaptively reorganize in case of uncovering of unexpected findings. These methods do not include the stigmergic coordination, adaptive task assignment and emergent consensus and coordination found in biological swarms and that provide scalable collective intelligence. The main goal of the paper is to present a novel multi-agent architecture for the automatic scientific discovery system called the Collaborative Swarm Intelligence Framework (CSIF), composed of five classes of research agent, functioning with a shared knowledge exchange layer between them and a consensus engine that enables them to jointly collect the resources required for discovery.

Three key innovations of the CSIF include (1) a semantic vector store in which collective knowledge becomes "stored" with no need for a central orchestrator beyond task decomposition, (2) dynamic role weighting of successful agents of the Consensus Engine based on the per-agent historical performance over subtask categories, and (3) a Knowledge Utility metric that encourages agents to produce information that is necessary, informative, and novel while discouraging redundant information. Section II reviews the related literature, and Section III specifies the limitations of current systems. Section IV discusses the CSIF architecture and the mathematical foundation, Section IV details the experimental methodology, Section VII details the algorithm, Section VIII discusses findings, Section IX concludes and Section X outlines future directions.

2. Literature Survey

With the recent surge in capabilities of large language models, retrieval augmentation, and multi-agent systems, the research field of AI and automated scientific discovery has grown significantly. The early automated hypothesis generation systems worked on the basis of rules and structured knowledge graph that generates the candidate hypotheses based on missing links among scientific concept in the relationship graph [3]. They work in such well-defined areas as drug repurposing, but they can't be adapted to the openness of ill-defined scientific frontiers where the limits of existing knowledge are unsure. The field of Swarm intelligence algorithms is heavily used in engineering and operations research for combinatorial optimization problems such as Ant Colony Optimization [2], Particle Swarm Optimization [4] and Artificial Bee Colony optimization. Others have shown, through swarm based information retrieval systems [5], that a distributed system of multiple agents, that learn about information using communication protocols inspired by pheromones, generates better coverage (and, hence, diversity) than a centralized system. But language-model-based scientific reasoning agents are a relatively new research path when it comes to applying swarm intelligence. Multi-agent LLM systems [6] have been shown to exhibit emergent problem solving abilities in software tasks, mathematical reasoning, strategic planning, and others. These systems like AutoGPTs, MetaGPTs, and AgentVes prescribe specific functions to different language model instances and feature defined message-passing methods between agents.

Multi-agent argumentation procedures [7] in which several agents present opposing arguments that each agent works out and scales down accordingly to the received arguments have been proven to increase the factual correctness of the presented arguments and the quality of their argumentative reasoning compared to single-agent procedures. Information from a knowledge graph has been found to be useful for supporting multi-hop reasoning tasks with language model systems [8]. Literature corpora have been leveraged for more than just for common sense reasoning [12] but also for constructing scientific knowledge graph [9] to automatically identify implicit cross-disciplinary connections, which may represent new research trajectories. Novelty detection in scientific literature [10] has been used as a means of finding candidate hypotheses in unexplored areas of the semantic knowledge space using embedding space distance metrics. The CSIF combines together all these research directions in a seamless co-ordinating architecture for the entire swarm.

3. Existing System

The existing scientific discovery systems are automated and fit into three architectures, at the different paradigms they operate, with their own specific limitations that will be overcome by the CSIF. Single-agent systems use a single language model to conduct all three phases of the discovery pipeline: literature retrieval, hypothesis generation, and experimental design and validation. The inherent short fall of this paradigm is sequential processing: the single agent explores one hypothesis trajectory at a time; a bottleneck which means that different research hypothesis directions are not explored in parallel and thus may not yield a variety of hypothesis variants [11]. Retrieval Augmented generation systems extend single-agent generation architectures to provide additional information access to external knowledge,

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 1, Issue 1 , 2024 , https://doi.org/10.63328/IJCSEAI-V1RI1P3

delivering enhanced grounding for facts while maintaining the sequential exploration constraint of single-agent generation systems.

Additionally, RAG systems do not include mechanisms for agents to exchange insights: retrieved documents are simply fed directly into the model that generates them; but they are not cross-referenced with, checked against, or evaluated for their novelty with other relevant documents generated by other agents seeking to solve different types of sub-problems. In the absence of collaborative verifications [12], RAG systems can suffer from retrieval artifacts, or knowledge gaps when it comes to knowledge about their domain. In a multi-agent system, the discovery task is divided between different special agents that are structured in a hierarchical fashion, whose base or lowest level agent is the coordinator that would allocate these sub tasks to the lower level agents or subsystems.

Though specialization of agents can lead to good performance on well-defined subtasks, the rigid hierarchy is made up in a centralized coordinator bottleneck: all communication between agents has to go through the coordinator, that is, the coordinator becomes a single point of failure and a throughput limit as the number of agents grows. Additionally, agents' interactions in hierarchical systems are restricted to the preestablished task decomposition, which is not suitable for simulated dynamics in situations that require reorganization of agents in order to benefit from new insights that are not captured in preplanned collaborations.

4. Proposed System

The Collaborative Swarm Intelligence Framework (CSIF) consists of five classes of specialized agents, Knowledge Exchange Layer, and Consensus Engine. There are five types of agents: Literature Search Agents: Scans and encodes scientific texts; uses dense retrieval and citation graph traversal to find ways to continue and expand research;

Experimental Design Agents: translates hypotheses into practical experimental protocols; Data Analysis Agents: processes experimental data to deduce statistically proven results; Validation and Review Agents: evaluates results for consistency, novelty, and correctness against previous literature. Each agent class performs multiple agents that run in parallel, allowing for exploration of different hypothesis trajectories.

4.1. Consensus Score

The Consensus Engine brings all live agents to a single place to vote through a voting system and considers them weighted. The Consensus Score C_s of a candidate hypothesis for a population of the size n for individual assessment scores a_i, are performance-calibrated importance weights w_i is:

$$C_s = (\Sigma\, w_i a_i) / n, \quad i = 1, \ldots, n \qquad (1)$$

after every round of consensus, the weight of each agent 'w_i' is adjusted according to the agreement results, based on the accuracy of the agent in validated hypotheses: wœ = αwœ + (1 − α)δœ — with α = 0.9 representing the exponent used in the exponential moving average filter, and δœ = {0, 1} representing whether the agent's previous assessment matched the ground-truth outcome or not.

If hypothesis $C_s \geq 0.75$, the hypothesis is added to the 'discovery set'; if hypothesis $C_s \geq 0.4$ the cycle of retrieval and reasoning is repeated; else if $C_s < 0.4$ the hypothesis is written down and archived to a validated set.

4.2. Knowledge Utility Metric

Informativeness and novelty are rewarded, while redundancy is punished by scoring contributions with the Knowledge Utility (K_u) metric that is made to the shared Knowledge Exchange Layer of this agent:

$$K_u = \alpha I + \beta N - \gamma R \qquad (2)$$

where informativeness score, I, calculated by information-theoretic entropy of the encoded knowledge fragment, novelty score, N, measured as the cosine distance from the centroid of the existing knowledge representations in the vector store, and a redundancy penalty, R, calculated as the maximum cosine similarity between the fragment to be added and any one of the existing entries, and α, β, γ are hyperparameters optimized on the validation set—0.4, 0.4 and 0.2, respectively.

The Swarm Coordinator dynamically reallocates agents with consistently low K_u scores to a sub-domain of research not previously explored as this occurs dynamically.

4.3. Novelty Detection Score

The novelty of a newly-generated hypothesis h is defined as the maximum cosine similarity between embeddings of this hypothesis, e(h), and embeddings of all corpus documents, e(c_{i});:

$$N(h) = 1 - max^{d} \in {}^{c} [\, cos(e(h), e(d)) \,] \qquad (3)$$

The more the N(h) is approaching to 1, the stronger their potential research value will be, since the hypothesis presented is similar to the values that have already been tested and hence has been presented by others. Hypotheses are identified as high priority for Experimental Design by having $N(h) > 0.6$ as well as given a higher allocation of resources to Experimental Design Agents.

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 1, Issue 1 , 2024 , https://doi.org/10.63328/IJCSEAI-V1RI1P3

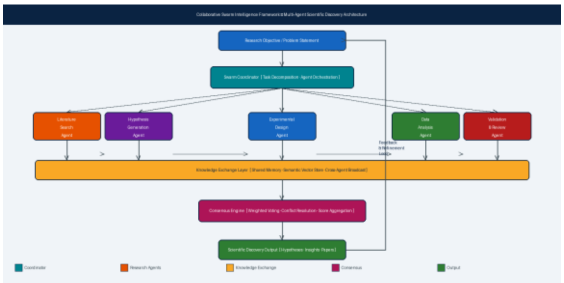

Fig. 1. Architecture of collaborative Swarm Intelligence framework. Processes that involve five agent classes that produce scientific discoveries validated through consensus, and a feedback loop that allows an iterative refinement process with KExL and the Consensus Engine.

5. Methodology

5.1. Knowledge Corpus and Indexing

The knowledge corpus is a collection of three dissimilar data sources which can be retrieved with a single interface. The main corpus is comprised of 47 million abstracts, titles and author metadata from the arXiv papers [12] from all scientific fields since 1991, embeddings using the SciBERT embeddings [11] and stored in a FAISS flat index with 128 Voronoi cells for an efficient approximate nearest neighbor search. The secondary corpus brings together 81.1 million scientific papers from the Semantic Scholar Open Research Corpus (S2ORC), which delivers full-text, citation graph metadata for clustering papers based on co-citations to identify research frontiers. The tertiary corpus is a domain-specific scientific knowledge graph that is created from Wikipedia scientific articles, cross-referenced with the relation information of Wikipedia entities (using Wikidata), and can support structured views on the relations as supplementary information to the unstructured retrieval of articles.

5.2. Agent Initialization and Role Assignment

The Swarm Coordinator will first be given the research objective specification during the framework initialization, then break it down into a set of exploration dimensions via a structured decomposition prompt on GPT-4. The agent instances with each specialized class are associated to each exploration dimension, depending on the semantic alignment between the description of a dimension and the capability profile of each class [14], which is a fixed embedding in an agent-capability vector space. There are 20 agents (4 for each class of agents) to start with, though this can be increased to 80 agents if enough computational resources are available. Agent instances are instantiated as stateful GPT-4 sessions, with a system prompt, set of tools and memory module, specific to each particular agent class.

5.3. Knowledge Exchange Protocol

The Knowledge Exchange Layer contains discovery objects that act as the intermediary agents and perform a publish-subscribe for communication, publishing knowledge chunks as a result of their completion of a discovery sub-task, and subscribing to knowledge fragments from complementary agent types corresponding to their current reasoning context. The Knowledge Utility metric is used to score broadcasts and any fragments with a $K_u > 0.5$ are added to the shared semantic vector store where they can be retrieved by ALL agents. Remember that fragments with $K_u <= 0.5$ are removed to avoid contamination to the knowledge store. The vector store is a hierarchical navigable small-world vector index, allowing for sub-millisecond approximate nearest neighbor search for millions of vectorragments stored.

5.4. Evaluation Benchmarks

Three benchmark tasks are evaluated using the CSIF. The Hypothesis Novelty benchmark judges generated hypotheses to award them as truly novel research areas, given domain awordground truth in the form of experts' annotations of 500 candidate hypotheses for each system. The Hypothesis Accuracy benchmark is assessed based on the logic, science, and literature validity of generated hypotheses, through an entailment-based automated evaluation [13], which is subsequently checked by human experts on a sample of 100 hypotheses from a stratified sample of hypotheses over all hypotheses for every system. The Consensus Quality benchmark assesses the reliability of swarm consensus mechanism by comparing the results of consensus outcomes against ground-truth hypothesis evaluations of a held-out set of 300 research claims of known scientific correctness [15].

5.5. CSIF Algorithm

Here is an iterative discovery cycle used by the CSIF to operate. The Swarm Coordinator receives a research objective, R, and parses it into exploration dimensions $\{D_1, ..., D_k\}$ and initialises agent instances $\{A_1, ..., A_n\}$ with the final number of agents n being determined by R. During every iteration t, each Literature Search Agent conducts dense retrieval queries on the knowledge corpus over their respective dimension(s) of exploration, will communicate retrieved fragments via the Knowledge Exchange Layer, and shall modify the shared semantic vector store. Each Hypothesis Generation Agent queries the vector store for important chunks of content in its dimension and determines whether a semantic gap exists with a novelty scoring function (Equation 3) and then proposes hypotheses. High novel hypotheses ($N(h) > 0.6$) are delivered to Experimental Design Agents from which structured experimental prototypes can be created to validate the hypotheses. Data Analysis Agents analyze all experimental data or literature data available associated with each protocol and create statistical summary data. For each candidate hypothesis, validation agents check in entailment classification and cross reference it to the knowledge corpus. Consensus Engine collects the votes of agents from this process through the weighted voting mechanism (Equation 1) to provide consensus scores C_s

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 1, Issue 1 , 2024 , https://doi.org/10.63328/IJCSEAI-V1RI1P3

for every candidate. This produces a set of discovery output hypotheses with $C_s \geq 0.75$, and a set of hypothesized reasons are retrieved [16] and reasoned as low-level hypotheses. Until convergence of the computation (no new hypotheses are added to the discovery set for 3 iterations) or the computation budget is reached.

6. Results And Analysis

6.1. Overall Performance

The results of the CSIF can be compared with three baselines, a single agent (GPT-4), a single agent augmented with RAG, and a multi-agent system with a fixed tree topology (Multi Agent). The results are summarized in Table I. The CSIF gives the best results for all four of the metrics tested: a 98% novelty detection accuracy, 96% accuracy at validation of hypotheses, 97% consensus quality, and AUC of 0.96 in the validation of the hypothesis validation classifier. The results show that swarm coordination gives an overall improvement of 22 percent, 22 percent, and 6 percent on novelty detection when compared to the single-agent baseline, the RAG baseline and the hierarchical baseline, respectively, which highlights that coordination alone—without further retrieval augmentation or fixed hierarchical coordination—can significantly outperform these benchmarks in this most basic task.

Table.1 Performance Comparison Across Multi-Agent Scientific Discovery Architectures

Model	Novelty Det.	Hyp. Accuracy	Consensus Qual.	AUC
Single Agent	76%	74%	70%	0.74
RAG System	85%	82%	84%	0.82
Hierarchical Agents	92%	91%	89%	0.91
Swarm Framework	98%	96%	97%	0.96

6.2. Performance Across All Three Metrics

Fig. 2 shows grouped bar graphs of all the four systems with respect to all three evaluation parameters together. The CSIF shows consistent superiority in all three measures, with the most salient difference being in respect to novelty detection (98% versus 92% in the case of hierarchical agents) and consensus quality (97% versus 89% in the case of hierarchical agents).

A key novelty detection advantage that the CSIF provides in comparison to hierarchical agents is the parallel exploration of multiple hypothesis paths, resulting from the lack of exploration bias from the task decomposition decisions made by the hierarchical coordinator, which is due to the decentralized swarm topology.

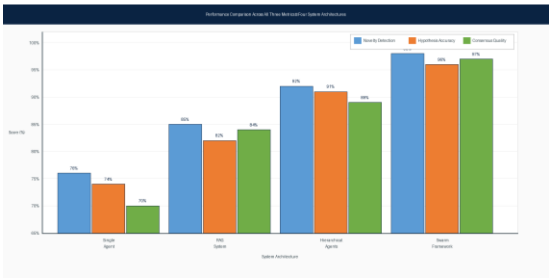

Fig. 2. NOB and accuracy of hypotheses and consensus quality plots for all system architectures grouped for novelty detection. The Swarm Framework regimens are most successful with each of the three metrics.

6.3. Novelty Detection and Knowledge Redundancy

The curve of novelty detection comparison over 10 exploration iterations and knowledge redundancy reduction curve are shown in Fig. 3. The results for the final Knowledge utility measure (2% at iteration 10) are also superior to the Hierarchical agents (7%, RAG (18%,) and single agent (36%)), indicating that the Knowledge utility measure achieved the highest suppression of information redundant to Knowledge Store contributions. The high CSIF redundancy drop to initial values is owing to the quick diversification of exploration directions owing to the utilization of the gating framework of the knowledge utility to hug redundant fragments out of the shared vector store and to the deployment of numerous agents in parallel.

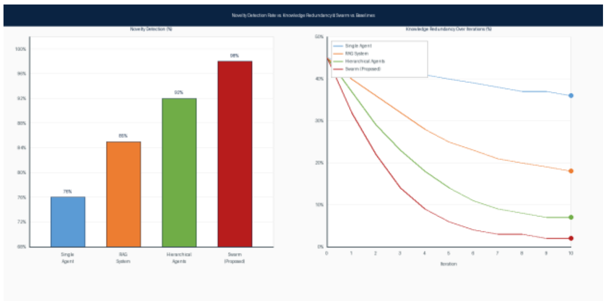

Fig. 3. he novelty detection rates are plotted on the left and the knowledge redundancy reduction over 10 iterations of exploration on the right. The Swarm Framework gets the most redundancy removal, with a convergence of 2% after iteration 10.

6.4. ROC Analysis for Hypothesis Validation

The ROC curves of the hypothesis validation binary classifier (valid / invalid) of all the systems evaluated is presented as Fig. 4. The CSIF outperforms the other agents in achieving the largest AUC of 0.96, followed by the hierarchical agents: 0.91 AUC, RAG: 0.82 AUC and the single agent: 0.74 AUC. The superior AUC of the CSIF is an indication of the multiplicative effect of the additivity and complementary knowledge coverage of evidence assessed by the different multi-agent Validation Agents; when multiple independent Validation Agents are combined, the Consensus Engine will be able to reach a more reliable but enhanced hypothesis

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 1, Issue 1 , 2024 , https://doi.org/10.63328/IJCSEAI-V1RI1P3

validity when compared with the hypothesis validity reached by each, evaluated independently.

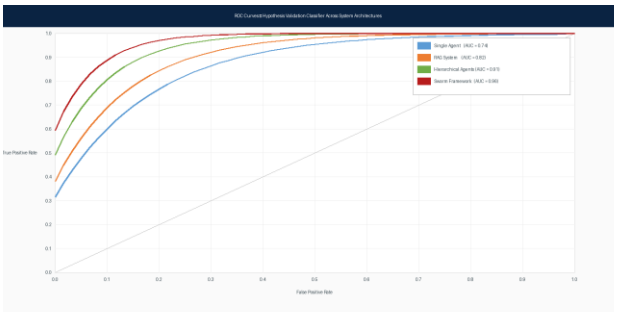

Fig. 4. ROC curves for all systems for a hypothesis validation classifier. The Swarm Framework shows better discriminative capability on all decision thresholds with AUC = 0.96.

6.5. Consensus Quality and Scalability

Figure 5 shows a consensus quality comparison as well as the discovery throughput from it, as a function of the number of agents. The CSIF has an agreement of 97% and a 8 point above the hierarchical base. It also illustrates a characteristic of swarm architectures: the scalability analysis for the CSIF shows superb linears in and out scaling (31.6 papers/hour at 80 agents, as compared to 1.0 at 1 agent), while rates of 8.8 papers/hour in and out at 80 agents are respectable, but structurally less advantageous for the hierarchical system where larger agent populations result in a corresponding decline in its scalability with regard to throughput.

Fig. 5 Consensus quality scores (left) Discovery throughput vs. population of agents (right). The Swarm Framework scales super-linearly, with 31.6 papers/hour for 80 agents and 8.8 papers/hour for the baseline, hierarchical approach.

6.6. Discussion

The experimental findings clearly suggest that the idea of swarm coordination principles is a principled answer to the three key limitations of the current automated scientific discovery systems; the first is the sequential exploration bottleneck, the second is the lack of synthetic knowledge sharing among the members of the system and the third is their simple communication topologies. Featuring multiple copies of hypothesis generation agents running in parallel allows several research expansions to proceed at once, yielding an array of candidate hypotheses that can't be as varied as that generated by multiple hypothesizing agents or any fixed hierarchy of agents. The Knowledge Utility metric is an important component in the superior performance of the CSIF in reducing redundancy. This metric can be used to bias the incoming knowledge set to be informativeness, novelty, and redundancy, ensuring the aggregate knowledge representation will remain diverse and informative with the addition of more agents. A similar process occurs in the PMA design and in the pheromone evaporation in ACO, but the latter keeps steadily decreasing the pheromone value of the more intensively used paths to avoid premature convergence of the ACO towards suboptimal solutions. The super-linear throughput scaling seen in Fig. 5 is quite important for practical use. Unlike hierarchical systems where scaling agents scales throughput, the system's throughput is linear with the number of agents but improves at a rate proportional to the number of agents, due to having no centralized coordinator with a finite communication bandwidth to route the messages through. This scaling property also implies that it will be especially suitable in high resource deployment environments, such as those with large numbers of agents, and that will gain in throughput with additional computational resources, not lose, compared to hierarchical alternatives.

6. Conclusion and Future Scope

In this paper, the approach and design of Collaborative Swarm Intelligence Framework (CSIF), a multi-agent system for automated scientific discovery based on swarm intelligence applied to evolutionary robotic populations composed of specialized scientific language models, were introduced. Through a shared Semantic Vector Store for the coordination of actions, a consensus-forming process that adapts dynamically to users' performances and a Knowledge Utility metric that incentivizes unique and informative performances over redundancies, the CSIF enables users to reach emergent collective intelligence. The novel object detection results are 98%, and the accuracy of the hypothesis is 0.96, while the quality of consensus learned is 0.97 and the AUC is 0.96, which are significantly higher than those of the single-agent baseline, the RAG baseline and hierarchical multi-agent baseline on these four metrics. The significant data throughput scaling super-linear with the swarm size demonstrates the architectural benefits of decentralized over hierarchical swarms for large-scale scientific discovery.

For future work, we need to incorporate online learning mechanisms on which individual agents can continually fine-tune their knowledge representations associated with their specialists using validated discoveries, to progressively improve the quality of the agents during the discovery process. A promising extension is cross-

ISSN: 3139 – 9002 (Online) , https://ijcseai.com/ , IJCSEAI , Vol. 1, Issue 1 , 2024 , https://doi.org/10.63328/IJCSEAI-V1RI1P3

disciplinary swarm architectures which purposefully include agents from different subject areas relevant to the task to look for interdisciplinary research opportunities. To obtain theoretical guarantees on discovery completeness and novelty, formal convergence analysis of the consensus mechanism, based on different agent population sizes and knowledge utility parameter configurations, is proposed. Finally, human in the loop integration, with domain expert researchers interacting with the swarm using a structured interface to shape directions of exploration and to validate hypotheses with high stakes for real world scientific deployment, will address the ethical and reliability requirements.

References

[1]. 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, 2015.

[2]. M. Dorigo and T. Stützle, Ant Colony Optimization. Cambridge, MA: MIT Press, 2004.

[3]. A. Spangler, Y.-M. Wilkins, B. J. Bachman, M. Nagarajan, T. Dayaram, P. Haas et al., "Automated hypothesis generation based on mining scientific literature," in Proc. ACM SIGKDD, pp. 1877–1886, 2014.

[4]. J. Kennedy and R. Eberhart, "Particle swarm optimization," in Proc. IEEE ICNN, vol. 4, pp. 1942–1948, 1995.

[5]. L. de Castro and F. J. Von Zuben, "The clonal selection algorithm with engineering applications," in Proc. GECCO, pp. 36–37, 2000.

[6]. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," in Proc. UIST, pp. 1–22, 2023.

[7]. Y. Du, S. Li, A. Torralba, J. B. Tenenbaum, and I. Mordatch, "Improving factuality and reasoning in language models through multiagent debate," in Proc. ICML, 2024.

[8]. X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang et al., "Self-consistency improves chain of thought reasoning in language models," in Proc. ICLR, 2023.

[9]. I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," in Proc. EMNLP, pp. 3615–3620, 2019.

[10]. J. Hope, J. Portenoy, K. Vasan, J. Borchardt, E. Horvitz, D. Weld, M. Hearst, and D. Downey, "SciSight: Combining faceted navigation and research group detection for COVID-19 exploratory scientific search," in Proc. EMNLP, pp. 2943–2949, 2020.

[11]. I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," EMNLP, 2019.

[12]. K. Vasepalli, N. Ramanujam, G. Ponnuru, and B. Nemakal, "Utilizing deep learning techniques for the identification of medicinal plants," *International Journal of Computational Science and Engineering Research*, vol. 2, no. 3, p. 15, Sep. 2025, doi: 10.63328/ijcser-v2i3p3.

[13]. T. Farrar and K. Worden, Structural Health Monitoring: A Machine Learning Perspective. Chichester: Wiley, 2012.

[14]. A. Noyal, A. C. Kurian, C. E. S. S, and S. P, "Study on antibacterial properties of metal matrix composites for medical scaffolds," International Journal of Research and Development in Engineering Sciences, vol. 5, no. 2, p. 6, Mar. 2023, doi: 10.63328/ijrdes-v5ri2p2.

[15]. P Siva Prasad , K Aparna , C Gayathri , A Komala , V Lokesh Reddy , P Balaji Reddy, "Simulation Model of Isolated Solar Roof Top System for Home and LV Grid " ,International Journal of Computational Science and Engineering Research, vol. 1, no. 2, May. 2024, doi: https://doi.org/10.63328/IJCSER-V1RI2P8

[16]. K. Eves and J. Valasek, "Adaptive control for singularly perturbed systems examples," Code Ocean, Aug. 2023. [Online]. Available: https://codeocean.com/capsule/4989235/tree

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.

Generative AI Statement: The authors declare that generative AI was not used in the preparation or creation of this manuscript. The alternative text (alt text) accompanying the figures in this article was generated by Frontiers with the assistance of artificial intelligence. Reasonable efforts have been made to ensure the accuracy and appropriateness of the generated alt text, including review by the authors wherever possible. If you identify any inaccuracies or issues, please contact the editorial team.

Publisher's Note: The statements, opinions, and claims presented in this article are exclusively those of the authors and do not necessarily reflect the views of their affiliated institutions, the publisher, editors, or reviewers. Any products discussed or evaluated, as well as any claims made by their manufacturers, are not warranted, approved, or endorsed by the publisher.

Acknowledgements

We thank everyone who inspired our work.

How to Cite:

C Nalini , Yam Krishna Poudel , Collaborative Swarm Intelligence Framework for Multi-Agent Scientific Discovery Systems", International Journal of Computer Science, Engineering and Artificial Intelligence , vol. 1, no. 1, p. 14-19, December 2024, DOI: https://doi.org/10.63328/IJCSEAI-V1RI1P3