

Neurosymbolic Framework for building Robust, Efficient, and Explainable Multi-Agent Enterprise Systems

Doctoral Consortium

Ritvik Garimella

University of South Carolina

Columbia, United States

ritvikg@sc.edu

ABSTRACT

Multi-agent systems for enterprise automation is an intelligent solution for a ubiquitous problem with three main objectives, robustness, efficiency and explainability. First, we study robustness in terms of automation procedures for similar tasks by aiming towards a construction of a hierarchical task decomposition with dynamic agent management system. Secondly, we study the concept of efficiency through reinforcement learning based algorithms for fine-tuned refinements of agent utilization and parametric optimization. Lastly, we investigate the concept of explainability by building evaluation suites for agents evaluation on both syntactic and semantic levels. The expected contributions include theoretical foundations for efficient orchestration and practical evaluation tools, advancing deployment of production-ready multi-agent systems.

KEYWORDS

Multi-Agent Systems; Autonomous Agents; Neurosymbolic AI; Multi-agent Orchestration; Multi-agent Evaluation

ACM Reference Format:

Ritvik Garimella. 2026. Neurosymbolic Framework for building Robust, Efficient, and Explainable Multi-Agent Enterprise Systems: Doctoral Consortium. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026)*, Paphos, Cyprus, May 25 – 29, 2026, IFAAMAS, 3 pages. <https://doi.org/10.65109/RXCC9483>

1 INTRODUCTION

Enterprise automation has evolved from Process Automation [4] through Robotic Process Automation (RPA) [1, 11], which introduced structured workflows but remained labor-intensive and brittle [5], to contemporary Agent Process Automation (APA) [12, 19, 21], where Large Language Models (LLMs) orchestrate workflows [2, 3, 15]. LLM-based multi-agent systems (LaMAS) [9, 10, 14, 20] are now getting used to solve complex tasks such as supply chain management [6], clinical decision support [18], and robotic coordination [16] by distributing specialized roles across multiple agents.

However, these LaMAS systems fundamentally coordinate homogeneous LLM-based agents through prompt engineering and role definitions, unable to orchestrate truly heterogeneous agent types, such as CNNs for vision, LSTMs for time-series analysis and more within the unified workflows. Recent advances introduce

trainable planning mechanisms [13], yet lack task oriented semantic grounding: the ability to understand task requirements, coordinate diverse computational agents, and verify outputs against domain constraints. Without comprehensive evaluation frameworks, these systems cannot ensure reliability in mission-critical applications such as manufacturing, healthcare, and autonomous systems.

My PhD research focuses on building a custom neurosymbolic orchestration where we integrate neural adaptability with symbolic reasoning to coordinate heterogeneous multi-agent systems [7]. In particular, we focus on developing a framework that decomposes high-level objectives into semantically coherent sub-tasks, constructs executable workflows coordinating heterogeneous agents, and enables continual learning through iterative refinement. The evaluation framework is designed to employ semantically-aware metrics for composite AI systems.

We analyze and investigate the problem domain across three interconnected dimensions:

- (1) neurosymbolic orchestration mechanisms for heterogeneous agent coordination.
- (2) comprehensive evaluation frameworks assessing orchestration quality at workflow and agent levels.
- (3) policy-based optimization for continual improvement of agent selection and workflow generation.

Our research has addressed the architectural mechanism for neurosymbolic orchestration and extends a community wide accepted evaluation metrics. We support our investigation through expert validations in manufacturing and finance domains. We are currently studying the creation of symbolic planners for task decompositions and integrating the first version of the evaluation suite, and plan to address the policy-optimization and expansion in future works.

2 MULTI-AGENT ORCHESTRATION

Orchestration Mechanisms. Current approaches coordinate homogeneous LLM-based agents through role differentiation and prompting strategies. AutoGen [20] enables programmatic multi-agent workflows, MetaGPT [9] assigns role-based specialization, and recent work introduces trainable planning. However, these systems cannot integrate heterogeneous computational agents within unified workflows, nor provide formal correctness guarantees. We developed a neurosymbolic orchestration framework integrating neural adaptability with symbolic reasoning. Task planning decomposes objectives through iterative loops where a neural planner proposes decompositions as directed acyclic graphs and a symbolic verifier validates against plan ontologies. Workflow orchestration constructs executable workflows by coordinating agents



This work is licensed under a Creative Commons Attribution International 4.0 License.

through constrained operators (sequential, branch, aggregate) and dual protocols for data-agent and inter-agent communication. Iterative improvement employs four feedback mechanisms: agent hyperparameter tuning, active learning over knowledge and data, workflow-level meta-learning, and ontology-guided plan refinement.

Evaluation Frameworks. Existing evaluation approaches focus on task completion and basic coordination metrics, lacking assessment of semantic correctness, interpretability, and domain alignment. Drawing from C3AN [17], we are developing a comprehensive evaluation suite assessing reliability, consistency, causal reasoning, grounding, and explainability at both agent and workflow levels. Comparative evaluations against AutoGen and AgentFlow across manufacturing and finance domains will identify which metrics predict deployment reliability.

Validation and Publications. We validated core components in smart manufacturing, coordinating time-series models, CNNs, and retrieval systems for multimodal anomaly detection. This work has been accepted for oral presentation at AAAI 2026 LaMAS Workshop [7] and a system demonstration at AAMAS 2026 [8].

3 RESEARCH OBJECTIVES

My PhD research aims to establish neurosymbolic orchestration as a foundational paradigm for reliable, interpretable multi-agent systems in mission-critical enterprise applications. This overarching goal encompasses three interconnected research objectives:

RQ1: How can neurosymbolic orchestration mechanisms coordinate heterogeneous agent types while ensuring semantic correctness? Current orchestration approaches handle homogeneous LLM-based agents but cannot integrate diverse computational components (neural networks for perception, symbolic reasoners for cognition, time-series models for temporal analysis) within verified workflows. This research objective investigates the design principles for orchestration frameworks that leverage symbolic reasoning to verify workflow correctness while enabling neural components to adapt to task requirements. Key challenges include defining compositional semantics for heterogeneous agent interactions, ensuring deterministic coordination without centralized communication, and scaling verification to complex enterprise workflows.

RQ2: What evaluation frameworks can comprehensively assess multi-agent orchestration quality across semantic, operational, and alignment dimensions? Existing metrics focus narrowly on task completion or basic coordination measures, failing to capture critical properties such as semantic grounding, interpretability, causal reasoning, and alignment with domain constraints. This objective develops a comprehensive evaluation suite that assesses orchestration at both workflow and agent levels, incorporating metrics for reliability, consistency, explainability, and safety. The challenge lies in operationalizing abstract properties like "semantic correctness" into measurable metrics and determining which evaluation dimensions most strongly predict system reliability in deployment.

RQ3: How can policy-based optimization enable continual improvement of orchestration decisions for agent selection and workflow generation? Multi-agent orchestration inherently

poses a multi-task learning challenge where the system must simultaneously learn effective task decomposition, optimal agent selection, and efficient workflow composition. This objective investigates reinforcement learning approaches for jointly optimizing these decisions through experience, developing mechanisms for workflow-level meta-learning and adaptive agent selection. Key questions include how to structure the optimization problem to balance exploration of new orchestration strategies with exploitation of proven patterns, and how to transfer learned orchestration knowledge across different enterprise domains.

These three objectives are deeply interconnected: effective orchestration mechanisms (RQ1) require robust evaluation to guide improvement (RQ2), while policy-based optimization (RQ3) depends on comprehensive metrics to assess orchestration quality. Together, they form a coherent research program toward trustworthy multi-agent systems for enterprise applications.

4 WORK PLAN

Phase 1: Neurosymbolic Orchestration Framework Developing three-stage architecture: (1) task planning with symbolic verification via iterative neural-symbolic loops, (2) workflow orchestration coordinating heterogeneous agents through constrained operators (sequential, branch, aggregate), and (3) iterative improvement infrastructure with four feedback mechanisms. Position paper on core orchestration design accepted at AAAI 2026 LaMAS Workshop; demonstration of agent coordination mechanisms under review at AAMAS 2026. Current work focuses on completing implementation of all orchestration components and expanding validation beyond initial smart manufacturing scenarios to establish generalizability across enterprise domains.

Phase 2: Comprehensive Evaluation Framework. Developing multi-dimensional evaluation suite operationalizing C3AN principles (14 principles including reliability, consistency, causal reasoning, explainability, etc.) into workflow-level and agent-level metrics. Implementing baseline measures, conducting comparative evaluations against AutoGen and AgentFlow across manufacturing and finance domains. Phase complete when evaluation library enables standardized assessment and identifies which metrics predict deployment reliability. Metrics guide Phase 3 optimization design.

Phase 3: Policy-Based Continual Improvement. Investigating policy gradient methods for joint optimization of task decomposition, agent selection, and workflow composition. Developing workflow-level meta-learning to transfer orchestration patterns across domains and centralized training with decentralized execution for agent coordination learning. Testing continual improvement mechanisms in manufacturing and process automation with Phase 2 metrics assessing impact on system reliability and interpretability.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Amit Sheth for his support and guidance.

REFERENCES

[1] Simone Agostinelli, Andrea Marrella, and Massimo Mecella. 2020. Towards intelligent robotic process automation for BPMers. *arXiv preprint arXiv:2001.00804* (2020).

[2] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. 2022. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691* (2022).

[3] Liying Cheng, Xingxuan Li, and Lidong Bing. 2023. Is gpt-4 a good data analyst? *arXiv preprint arXiv:2305.15038* (2023).

[4] Andrzej Cichocki, Helal A Ansari, Marek Rusinkiewicz, and Darrell Woelk. 1997. *Workflow and process automation: concepts and technology*. Vol. 432. Springer Science & Business Media.

[5] Marc Eulerich, Nathan Waddoups, Martin Wagener, and David A. Wood. 2024. The Dark Side of Robotic Process Automation (RPA): Understanding Risks and Challenges with RPA. *Accounting Horizons* 38, 2 (06 2024), 143–152. <https://doi.org/10.2308/HORIZONS-2022-019> arXiv:<https://publications.aaahq.org/accounting-horizons/article-pdf/38/2/143/108473/i0888-7993-38-2-143.pdf>

[6] Md Zahidur Rahman Farazi. 2025. Enhancing supply chain resilience with multi-agent systems and machine learning: a framework for adaptive decision-making. *The American Journal of Engineering and Technology* 7, 03 (2025), 6–20.

[7] Ritvik Garimella, Chathurangi Shyalika, Renjith Prasad Kaippilly Mana, and Amit Sheth. 2026. DYN0: Dynamic Neurosymbolic Orchestrator for Multi-Agent Systems. In *LLM-based Multi-Agent Systems: Towards Responsible, Reliable, and Scalable Agentic Systems*.

[8] Ritvik Garimella, Michael Charles Stewart, Kevin Do, Leeon Israel, and Amit Sheth. 2026. KiMO: Knowledge-infused Multi-agent Orchestrator. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Demonstration Track*. <https://doi.org/10.65109/WTIF5096> Accepted for publication.

[9] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, et al. 2023. MetaGPT: Meta programming for a multi-agent collaborative framework. In *The Twelfth International Conference on Learning Representations*.

[10] Mengkang Hu, Yuhang Zhou, Wendong Fan, Yuzhou Nie, Bowei Xia, Tao Sun, Ziyu Ye, Zhaoxuan Jin, Yingru Li, Qiguang Chen, et al. 2025. Owl: Optimized workforce learning for general multi-agent assistance in real-world task automation. *arXiv preprint arXiv:2505.23885* (2025).

[11] Lucija Ivančić, Dalia Suša Vugec, and Vesna Bosilj Vuksić. 2019. Robotic process automation: systematic literature review. In *International Conference on Business Process Management*. Springer, 280–295.

[12] Zelong Li, Shuyuan Xu, Kai Mei, Wenyue Hua, Balaji Rama, Om Raheja, Hao Wang, He Zhu, and Yongfeng Zhang. 2024. Autoflow: Automated workflow generation for large language model agents. *arXiv preprint arXiv:2407.12821* (2024).

[13] Zhuofeng Li, Haoxiang Zhang, Seungju Han, Sheng Liu, Jianwen Xie, Yu Zhang, Yejin Choi, James Zou, and Pan Lu. 2025. In-the-Flow Agentic System Optimization for Effective Planning and Tool Use. <https://doi.org/10.48550/arXiv.2510.05592> arXiv:2510.05592 [cs.AI]

[14] OpenAI. 2025. API Agents | OpenAI. <https://openai.com/agent-platform/>. Accessed: 2025-11-02.

[15] Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, et al. 2023. Chatdev: Communicative agents for software development. *arXiv preprint arXiv:2307.07924* (2023).

[16] Jiayuan Rao, Zifeng Li, Haoning Wu, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2025. Multi-agent system for comprehensive soccer understanding. *arXiv preprint arXiv:2505.03735* (2025).

[17] Amit P. Sheth, Kaushik Roy, Revathy Venkataramanan, Venkatesan Nadimuthu, and Chathurangi Shyalika. 2025. Composite AI With Custom, Compact, Neurosymbolic Models: The Emergent Enterprise Artificial Intelligence Paradigm. *IEEE Internet Computing* 29, 2 (2025), 37–49. <https://doi.org/10.1109/MIC.2025.3570554>

[18] Andre Lehdermann Silveira, Rodrigo da Rosa Righi, and Cristiano Andre da Costa. 2025. Multi-agent systems for clinical decision support: A systematic review. *Applied Soft Computing* (2025), 114447.

[19] Michael Wornow, Avanika Narayan, Krista Opsahl-Ong, Quinn McIntyre, Nigam H Shah, and Christopher Re. 2024. Automating the enterprise with foundation models. *arXiv preprint arXiv:2405.03710* (2024).

[20] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, et al. 2024. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First Conference on Language Modeling*.

[21] Yining Ye, Xin Cong, Shizuo Tian, Jiannan Cao, Hao Wang, Yujia Qin, Yaxi Lu, Heyang Yu, Huadong Wang, Yankai Lin, et al. 2023. Proagent: From robotic process automation to agentic process automation. *arXiv preprint arXiv:2311.10751* (2023).