Project Proposal: Scalable Multi-Agent System

Andrea Grandi
Master's Degree in Artificial Intelligence Engineering, UniMoRe

October 9, 2025

Abstract

This project aims to design and evaluate a scalable multi-agent system built on top of the LangGraph framework. The main objective is to study how the system behaves and scales with respect to (1) the number of agents, (2) the size of the individual agents, and (3) the architecture of the multi-agent setup. Different organizational structures-single-agent, supervisor-based, and swarm (peer-to-peer) will be implemented and compared through controlled experiments using established benchmarks. The project also explores the economic trade-offs of using smaller specialized agents and the explainability of multi-agent interactions.

Keywords: Multi-Agent Systems, LangGraph, Scalability, AgentBench, MultiAgentBench, AgentForest, Explainability, Distributed AI

1 Background and Motivation

Multi-agent systems (MAS) are increasingly used to model distributed intelligent systems, where autonomous entities cooperate or compete to achieve complex goals. Recent frameworks such as **LangGraph** enable composable and structured agentic architectures, offering a foundation for developing and analyzing intelligent agent networks. However, while multi-agent coordination has been widely explored, systematic studies on **scalability** how performance, cost, and interpretability evolve with the number and complexity of agents are still limited.

In recent literature, various challenges in LLM-based multi-agent systems are discussed, including task allocation, memory alignment, and coordination complexity [3], as well as the emerging view that **Small Language Models** (**SLMs**) (Belcak et al. 2025) [2] may be more economical and scalable in agentic systems. Work on heterogeneous agent systems also underscores that mixing different agent sizes/types can yield efficiency gains [1]. Leveraging **Model Context Protocol** (**MCP**) (Xinyi et al., 2025) [4] in large-scale systems also pose additional constraints. Finally, models for scalable agent swarm provide useful methodological inspiration for benchmarking and system scaling studies like **AgentForest** (Li et al., 2025) [5].

This thesis aims to build on these foundations and conduct an empirical, comparative study of scalable MAS architectures, focusing on performance, economic cost, and explainability.

2 Objectives

The project has the following main goals:

- **Architecture Design:** Develop three types of agentic architectures using LangGraph:
  1. Single Agent Architecture
  2. Multi-Agent Orchestrator (Supervisor-based)
  3. Swarm Architecture (Peer-to-Peer)
- **Scalability Analysis:** Investigate how system performance changes with:
  - The number of agents
  - The model size and capability of individual agents
  - The type of multi-agent architecture
- **Economic and Interpretability Evaluation:**
  - Assess the cost-efficiency trade-offs of smaller, specialized agents versus large monolithic models
  - Study the explainability and interpretability of multi-agent decision-making

3 Methodology

3.1 Framework

The system will be developed using the **LangGraph** framework as the core for defining agent behaviors, tools, and communication graphs.

3.2 Architectures

1. **Single Agent:** A single autonomous agent executing tasks independently.
2. **Supervisor-based Multi-Agent System:** A hierarchical architecture with an orchestrator coordinating several specialized sub-agents.
3. **Swarm Architecture:** A peer-to-peer model where all agents communicate and collaborate without a central coordinator.

3.3 Benchmarks and Experiments

The following benchmark frameworks will be employed:

- **Single-Agent Evaluation:** Using **AgentBench** (Liu et al., 2023) [6] (`https://github.com/THUDM/AgentBench`) to test individual agents on OS or database tasks.
- **Architecture Evaluation:** Using **MultiAgentBench** (`https://github.com/hinthornw/mabench`) to evaluate coordination efficiency and performance across architectures.
- **System Scalability:** Using methods inspired by **AgentForest** (`https://github.com/MoreAgentsIsAllYouNeed/AgentForest`) and **MARBLE** (Zhu et al., 2025) [7] (`https://github.com/ulab-uiuc/MARBLE`) to study system scalability as the number of agents grows.

3.4 Evaluation Metrics

- **Agent-Level:** Accuracy, efficiency, robustness as a function of model size.
- **Architecture-Level:** Coordination cost, task completion rate, latency, communication overhead.
- **System-Level:** Scalability, stability, throughput, emergent behavior under increasing populations.
- **Economic Evaluation:** Cost vs performance trade-offs between large and small agents.
- **Explainability:** Qualitative and quantitative assessment of interpretability in agent reasoning and interactions.

4 Expected Results

The project is expected to deliver:

- A modular, extensible implementation of a scalable multi-agent system built on LangGraph.
- Quantitative analysis of system scalability, performance, and cost efficiency.
- Insights into trade-offs between agent model size, architecture design, and system scalability.
- Comparative evaluation of explainability across agentic architectures.

5 Conclusion

This thesis aims to contribute to the field of multi-agent systems by offering an in-depth empirical study on scalability, cost, and design trade-offs. Leveraging LangGraph as a flexible foundation, the project seeks to derive practical guidelines for building efficient, interpretable, and scalable multi-agent systems.

References

[1] Zain Asgar, Michelle Nguyen, and Sachin Katti. Efficient and scalable agentic ai with heterogeneous systems, 2025.

[2] Peter Belcak, Greg Heinrich, Shizhe Diao, Yonggan Fu, Xin Dong, Saurav Muralidharan, Yingyan Celine Lin, and Pavlo Molchanov. Small language models are the future of agentic ai, 2025.

[3] Shanshan Han, Qifan Zhang, Yuhang Yao, Weizhao Jin, and Zhaozhuo Xu. Llm multi-agent systems: Challenges and open problems, 2025.

[4] Xinyi Hou, Yanjie Zhao, Shenao Wang, and Haoyu Wang. Model context protocol (mcp): Landscape, security threats, and future research directions, 2025.

[5] Junyou Li, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye. More agents is all you need, 2024.

[6] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang. Agentbench: Evaluating llms as agents. *arXiv preprint arXiv: 2308.03688*, 2023.

[7] Kunlun Zhu, Hongyi Du, Zhaochen Hong, Xiaocheng Yang, Shuyi Guo, Zhe Wang, Zhenhailong Wang, Cheng Qian, Xiangru Tang, Heng Ji, and Jiaxuan You. Multiagentbench: Evaluating the collaboration and competition of llm agents, 2025.