

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Thesis Proposal

**Design and Evaluation of Multi-Agent AI Systems for
Contract Lifecycle Management in B2B SaaS Platforms**

| | |
|------------------|-----------------------------|
| Author: | Trung Nguyen |
| Supervisor: | Prof. Dr. [Supervisor Name] |
| Advisor: | [Advisor Name] |
| Submission Date: | October 8, 2025 |

Motivation

In recent years, Contract Lifecycle Management (CLM) [Wik25] has increasingly adopted artificial intelligence to automate tasks such as clause extraction, risk scoring, and contract summarization. These advancements demonstrate the potential of AI-driven CLM systems to reduce manual workload and accelerate contract turnaround times. However, current solutions typically rely on monolithic, single-agent large language models (LLMs) that offer limited explainability, weak controllability, high hallucination risk [BÇ25], and poor scalability across diverse contract types and jurisdictions [Bil+23; Xia+25]. Such black-box architectures make it difficult to ensure consistency, traceability, and compliance—capabilities that are critical in regulated, multi-stakeholder enterprise environments. [GR21; Wan+24a]

Multi-agent systems (MAS) have recently emerged as a promising paradigm for addressing these limitations. By decomposing contract workflows into specialized, collaborating agents, MAS can enable modular scalability, explicit reasoning chains, and improved explainability [Shu+24; Wan+24b]. Despite this potential, their application to enterprise CLM remains largely unexplored. Existing studies highlight a clear research gap: the absence of frameworks and empirical evaluation for compliance-aware, explainable multi-agent orchestration in enterprise contract management [Yeh+25].

Research Questions

How can a multi-agent framework be designed and implemented to improve contract analysis efficiency, accuracy, and explainability in enterprise CLM systems?

Research Objectives:

1. Design a focused multi-agent framework for contract analysis based on industry foundation papers
2. Develop a working prototype with 2-3 specialized agents
3. Conduct comparative evaluation using real contract data
4. Establish practical implementation guidelines for enterprise CLM integration

Methodology

Foundation Framework Design Based on industry foundation papers (e.g., Wooldridge's MAS principles, enterprise AI frameworks), design a focused multi-agent architecture for contract analysis. Define agent roles: Contract Analyzer Agent (extraction, classification), Compliance Checker Agent (regulatory validation), and Orchestrator Agent (task coordination).

Prototype Implementation Develop a working prototype implementing the designed framework. Core components: (1) Agent communication protocols using message passing, (2) Contract processing pipeline with NLP integration, (3) Basic explainability features (decision reasoning, confidence scores), (4) Simple web interface for testing.

Evaluation and Validation Conduct comparative evaluation using real contract datasets. Metrics: (1) Processing accuracy (precision, recall for contract analysis tasks), (2) Performance comparison (multi-agent vs single-agent processing time), (3) User trust assessment through explainability features, (4) System reliability and error handling.

Implementation Approach: Rapid prototyping using Python with existing NLP libraries (spaCy, transformers), agent framework (asyncio for concurrent processing), and evaluation using standard contract analysis benchmarks.

Schedule

The research project spans from October 2025 to March 2026, structured into three phases with dedicated time for thesis writing.

Phase 1: Literature Review + Detailed Planning (October - November 2025)

- Study multi-agent systems and enterprise AI frameworks
- Design multi-agent architecture for contract analysis (2-3 specialized agents)
- Define agent roles, communication protocols, and system requirements

Phase 2: Prototype Implementation (December 2025 - January 2026)

- Develop working prototype using relevant agentic frameworks and libraries.
- Implement modular architecture with JSON/YAML configuration files
- Establish agent reasoning pipeline and basic explainability features
- **Key:** Avoid hard-coded features; use configurable parameters for easy experimentation

Phase 3: Evaluation and Validation (February - March 2026)

- Conduct comparative evaluation using real contract datasets
- Test multi-agent system against single-agent baseline
- **Thesis Writing:** Reserve final 3-4 weeks for writing and quick fixes

Milestones

- **November 2025:** Framework design complete, agent roles defined
- **January 2026:** Working prototype with core functionality achieved
- **March 2026:** Evaluation complete, thesis writing finished

Expected Contributions

- Working prototype implementation with documented architecture, agent design patterns, and integration approach for enterprise CLM systems
- Evaluation framework and metrics for assessing multi-agent contract analysis systems, with practical guidelines for implementation
- Practical explainability features for enterprise contract analysis, validated through user testing and trust assessment

Impact Areas: Contract analysis automation, Multi-Agent Systems in enterprise software, Legal Technology prototyping, and practical AI implementation patterns.

Bibliography

- [BÇ25] Y. Balı and B. Çiloğlu. “Leveraging Large Language Models for Natural Language Processing Based Tasks in the Legal Domain: A Short Survey.” In: *Computational Science and Its Applications – ICCSA 2025: 25th International Conference, Istanbul, Turkey, June 30–July 3, 2025, Proceedings, Part II*. Istanbul, Türkiye: Springer-Verlag, 2025, pp. 166–178. ISBN: 978-3-031-96996-6. DOI: 10.1007/978-3-031-96997-3_11.
- [Bil+23] M. Billi, A. Parenti, G. Pisano, and M. Sanchi. *Large Language Models and Explainable Law: a Hybrid Methodology*. 2023. arXiv: 2311.11811 [cs.AI].
- [GR21] J. Gryz and M. Rojszczak. “Black box algorithms and the rights of individuals: No easy solution to the “explainability” problem.” In: *Internet Policy Review* 10 (June 2021). DOI: 10.14763/2021.2.1564.
- [Shu+24] R. Shu, N. Das, M. Yuan, M. Sunkara, and Y. Zhang. *Towards Effective GenAI Multi-Agent Collaboration: Design and Evaluation for Enterprise Applications*. 2024. arXiv: 2412.05449 [cs.CL].
- [Wan+24a] J. Wang, H. Zhao, Z. Yang, P. Shu, J. Chen, H. Sun, R. Liang, S. Li, P. Shi, L. Ma, Z. Liu, Z. Liu, T. Zhong, Y. Zhang, C. Ma, X. Zhang, T. Zhang, T. Ding, Y. Ren, T. Liu, X. Jiang, and S. Zhang. *Legal Evaluations and Challenges of Large Language Models*. 2024. arXiv: 2411.10137 [cs.CL].
- [Wan+24b] J. Wang, J. Wang, B. Athiwaratkun, C. Zhang, and J. Zou. *Mixture-of-Agents Enhances Large Language Model Capabilities*. 2024. arXiv: 2406.04692 [cs.CL].
- [Wik25] Wikipedia contributors. *Contract lifecycle management*. Last edited on 5 September 2025, at 06:44 (UTC). 2025.
- [Xia+25] C. Xia, Q. Wu, S. Tian, and Y. Hao. *Parallelism Meets Adaptiveness: Scalable Documents Understanding in Multi-Agent LLM Systems*. 2025. arXiv: 2507.17061 [cs.MA].
- [Yeh+25] A. Yehudai, L. Chen, M. Rodriguez, and S. Kim. *LLM Agent Evaluation in Regulated Enterprise Domains: A Comprehensive Survey*. 2025. arXiv: 2503.16416 [cs.AI].