DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Thesis Proposal

Design and Evaluation of Multi-Agentic 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; WY25]. 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; Kar25].

Research Questions

How can a multi-agent framework be designed and implemented to improve contract analysis efficiency, accuracy, and explainability in enterprise CLM systems, which could be broken down into the following sub-questions:

1. What agent specialization patterns are most effective for contract analysis tasks (drafting, review, compliance checking)?

2. How does a multi-agent approach compare to single-agent systems in contract analysis accuracy and processing time?

3. What explainability mechanisms are most effective for multi-agent contract analysis systems to ensure transparency and trust in enterprise environments?

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 Grounded in domain expertise from enterprise and construction contract management, this research identifies four key agent roles that address practical challenges in Contract Lifecycle Management (CLM): (1) *Risk Analysis Agent* – detects red-flag clauses, liability gaps, and unbalanced terms during contract drafting; (2) *Clause Alignment Agent* – ensures consistency across interrelated project contracts by recommending standardized or back-to-back clauses from the clause library; (3) *Obligation Tracking Agent* – monitors post-signature obligations and deadlines across stakeholders to improve compliance; and (4) *Dependency Graph Agent* (forward-looking) – maintains a dynamic representation of clause and obligation relationships across multiple contracts. These roles were derived through consultation with domain experts and analysis of current CLM pain points in large construction projects. The multi-agent architecture will be designed following established MAS principles (e.g., Wooldridge's frameworks) and enterprise AI orchestration patterns, focusing on modularity, explainability, and compliance awareness.

Prototype Implementation A working prototype will operationalize the designed architecture using modern agentic AI frameworks. Core components include: (1) Inter-agent communication and orchestration implemented through *LangGraph*, enabling concurrent, event-driven collaboration among specialized agents; (2) A contract processing pipeline leveraging *AWS Textract and custom OCR pipeline* for document parsing, clause extraction, and retrieval-augmented reasoning using vector embeddings stored in *Supabase's vector database*; (3) Built-in explainability and observability through *LangFuse*, providing detailed reasoning traces, confidence scores, and user-level audit logs; and (4) A lightweight web interface for interactive testing of agent behavior and real-time contract analysis scenarios. The prototype will be developed in Python, integrating existing LLM APIs (e.g., GPT-4 or similar models) for natural language understanding and leveraging asynchronous orchestration within the LangGraph framework to simulate enterprise-grade agent workflows.

Evaluation and Validation The evaluation will assess the proposed multi-agent framework across three dimensions:

1. **Efficiency and Accuracy** – Comparative study against single-agent LLM baseline using precision, recall, and F1 metrics for risk detection, and overall processing time for efficiency measurement

2. **Consistency and Compliance** – Testing how well the system maintains consistency across related contracts and accurately tracks obligations and deadlines

3. **Explainability and Trust** – Qualitative assessment using reasoning-trace coverage, auditability through LangFuse logs, and expert Likert-scale ratings of clarity and reliability

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. Çiloğlugil. "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.

[Kar25] A. Kargwal. *Mastering Multi-Agent Eval Systems in 2025*. Blog post. Botpress Blog. Jan. 2025.

[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 Evalutions 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.

[WY25]   Z. Wang and B. Yuan. *L-MARS: Legal Multi-Agent Workflow with Orchestrated Reasoning and Agentic Search*. 2025. arXiv: 2509.00761 [cs.AI].

[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].