DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

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

Trung Nguyen

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

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

Design und Evaluierung von Multi-Agenten KI-Systemen für Contract Lifecycle Management in B2B SaaS Plattformen

Author:	Trung Nguyen
Supervisor:	Prof. Dr. [Supervisor Name]
Advisor:	[Advisor Name]
Submission Date:	[Submission Date]

I confirm that this master's thesis in informatics is my own work and I have documented all sources and material used.

Munich, [Submission Date] Trung Nguyen

Acknowledgments

Abstract

Contract Lifecycle Management (CLM) is critical in project-based industries such as construction, energy, and telecommunications. Current CLM tools rely on rule-based automation or single-agent Large Language Models (LLMs), resulting in limited scalability, poor explainability, and lack of domain-specific optimization. This thesis proposes the design and evaluation of multi-agentic AI systems for contract lifecycle management in B2B SaaS platforms.

The research addresses key challenges in current CLM systems: lack of scalable orchestration for complex workflows, explainability gaps in AI-driven decisions, and insufficient compliance and regulatory constraint handling. The proposed approach leverages multi-agent systems (MAS) to enable specialized task decomposition, collaboration, and enhanced explainability in enterprise workflows.

This work contributes to both academic research and practical applications by developing a framework for MAS design in regulated enterprise workflows, providing comparative empirical results between single-agent and multi-agent approaches, and establishing evaluation methodologies for compliance-aware AI systems. The research directly informs the development of next-generation CLM platforms while maintaining discretion regarding proprietary implementation details.

Keywords: Multi-agent systems, Contract Lifecycle Management, Enterprise AI, Compliance-aware systems, B2B SaaS

Contents

1 Introduction

1.1 Context and Motivation

Contract Lifecycle Management (CLM) is critical in project-based industries such as construction, energy, and telecommunications. In these sectors, contracts serve not merely as legal documents but as operational and financial anchors that govern complex, long-term projects involving multiple stakeholders, regulatory requirements, and dynamic risk factors.

Current CLM tools predominantly rely on rule-based automation or single-agent Large Language Models (LLMs), which present significant limitations in scalability, explainability, and domain-specific optimization. These approaches struggle to handle the complexity of enterprise workflows that require specialized task decomposition, collaborative decision-making, and transparent reasoning processes.

Multi-agent systems (MAS) offer promising potential for addressing these limitations through specialized task decomposition, inter-agent collaboration, and enhanced explainability mechanisms. The distributed nature of MAS aligns well with the inherently collaborative and multi-stakeholder environment of contract management processes.

1.2 Research Relevance

This research addresses both academic and practical needs:

Academic Relevance: The work fills a research gap on multi-agent system applications in enterprise workflows within regulated environments. While MAS have been extensively studied in domains such as robotics and conversational AI, their application to enterprise contract management remains underexplored, particularly regarding compliance-aware orchestration and explainability requirements.

Practical Relevance: The research contributes directly to the development of next-generation CLM platforms. By focusing on methodology and evaluation rather than proprietary implementation details, this work provides valuable insights for the broader enterprise software community while maintaining discretion regarding sensitive commercial aspects.

1.3 Scope and Limitations

This thesis focuses on the design and evaluation of multi-agentic AI systems specifically for contract lifecycle management tasks. The scope encompasses:

- System architecture design for MAS-based CLM workflows
- Orchestration and collaboration strategies for enterprise environments
- Comparative evaluation against single-agent baselines
- User trust and explainability assessment

The research maintains discretion regarding proprietary implementation details while providing generalizable frameworks and methodologies applicable to the broader enterprise AI domain.

2 Problem Statement

2.1 Challenges in Current CLM Systems

Contract Lifecycle Management systems face several critical challenges that limit their effectiveness in enterprise environments:

2.1.1 Scalability and Orchestration Limitations

Current CLM tools struggle with scalable orchestration for complex workflows involving drafting, negotiation, and compliance monitoring. Traditional rule-based systems require extensive manual configuration and maintenance, while single-agent LLM approaches lack the modularity needed for specialized task handling across different contract types and regulatory domains.

2.1.2 Explainability Gap

AI-driven decisions in contract management often lack sufficient explainability, leading to low enterprise trust and adoption barriers. Users need to understand not only what decisions are made but also the reasoning process behind contract clause recommendations, compliance assessments, and risk evaluations.

2.1.3 Compliance and Regulatory Constraints

Existing AI-based CLM systems inadequately address compliance and regulatory constraints that vary across industries and jurisdictions. The dynamic nature of regulatory requirements and the need for audit trails present significant challenges for current approaches.

2.2 Research Gap

The current state of research reveals several critical gaps:

- **No established frameworks** for multi-agent orchestration specifically designed for enterprise CLM workflows

- **Limited empirical evaluation** comparing single-agent versus multi-agent approaches in contract management contexts
- **Insufficient attention** to compliance-aware AI system design in enterprise environments
- **Lack of standardized evaluation methodologies** for assessing trust and explainability in AI-driven contract management

2.3 Core Research Problem

The central problem addressed by this research is: **How can multi-agent systems be designed and evaluated to improve efficiency, accuracy, and trust in contract lifecycle processes?**

This problem encompasses three key dimensions:

1. **Design Challenge:** Developing appropriate MAS architectures for CLM workflows
2. **Coordination Challenge:** Establishing effective orchestration and collaboration strategies
3. **Evaluation Challenge:** Creating comprehensive assessment frameworks for performance, compliance, and user trust

The solution to this problem requires interdisciplinary research combining multi-agent systems theory, enterprise software architecture, and human-computer interaction principles to create practical, trustworthy, and scalable CLM solutions.

3 Research Questions

3.1 Primary Research Question

The central research question driving this investigation is:

How can multi-agent systems be designed and evaluated to improve efficiency, accuracy, and trust in contract lifecycle management processes?

This primary question encompasses the design, implementation, and evaluation aspects of MAS-based CLM systems, focusing on practical applicability in enterprise environments.

3.2 Secondary Research Questions

To address the primary research question comprehensively, this thesis investigates three secondary research questions:

3.2.1 RQ1: System Design

How can multi-agent systems be architected for contract lifecycle management tasks such as clause extraction, drafting assistance, and compliance monitoring?

This question focuses on the technical architecture and design patterns suitable for MAS-based CLM systems. It explores:

- Agent specialization strategies for different CLM tasks
- Communication protocols and message passing mechanisms
- System scalability and modularity considerations
- Integration with existing enterprise infrastructure

3.2.2 RQ2: Coordination and Orchestration

What orchestration and collaboration strategies enable efficiency and explainability in multi-agent CLM systems?

This question addresses the coordination mechanisms that ensure effective collaboration between agents while maintaining system transparency. Key aspects include:

- Task decomposition and assignment strategies
- Conflict resolution mechanisms
- Explainability frameworks for multi-agent decisions
- Performance optimization techniques

3.2.3 RQ3: Comparative Evaluation

How does a multi-agent approach compare to a single-agent baseline on performance, compliance adherence, and user trust?

This question establishes empirical evidence for the effectiveness of MAS approaches through systematic comparison. Evaluation dimensions include:

- Task completion accuracy and efficiency
- Compliance adherence rates
- User trust and acceptance metrics
- System reliability and robustness

3.3 Research Hypotheses

Based on the research questions, the following hypotheses guide the investigation:

1. **H1:** Multi-agent systems will demonstrate superior task specialization and modularity compared to single-agent approaches in CLM workflows.
2. **H2:** MAS-based CLM systems will provide enhanced explainability through distributed reasoning and specialized agent roles.
3. **H3:** Multi-agent approaches will achieve higher compliance adherence rates due to specialized compliance monitoring agents.
4. **H4:** Users will demonstrate higher trust in MAS-based CLM systems due to improved transparency and explainability.

3.4 Success Criteria

The success of this research will be measured through the following criteria:

- **Technical Performance:** Demonstrable improvements in task completion accuracy and efficiency compared to baseline systems
- **Compliance Achievement:** Measurable enhancement in regulatory compliance adherence
- **User Acceptance:** Positive user feedback regarding trust, usability, and explainability
- **System Scalability:** Evidence of system performance under varying workload conditions
- **Research Contribution:** Publication-quality results that advance the state of knowledge in MAS and enterprise AI

4 Methodology

4.1 Literature Review

A comprehensive literature review will establish the theoretical foundation for this research, covering three primary domains:

4.1.1 Multi-Agent Orchestration

Review of existing MAS orchestration frameworks, focusing on:

- Coordination mechanisms in conversational AI and robotics
- Task decomposition and assignment strategies
- Communication protocols and message passing systems
- Scalability and performance optimization techniques

4.1.2 AI in Legal Technology and Contract Management

Analysis of current applications of AI in legal tech, including:

- Natural language processing for contract analysis
- Machine learning approaches to contract classification
- Automated contract generation and review systems
- Compliance monitoring and risk assessment tools

4.1.3 Compliance-Aware and Explainable AI Frameworks

Examination of frameworks for building trustworthy AI systems:

- Explainable AI methodologies and techniques
- Compliance-aware system design principles
- Trust and transparency in enterprise AI systems
- Evaluation methodologies for AI system reliability

4.2 Prototype Development

A functional MAS-based CLM prototype will be developed to validate the proposed approach. The prototype will focus on core CLM tasks while maintaining discretion regarding proprietary implementation details.

4.2.1 System Architecture

The prototype will implement:

- Specialized agents for different CLM tasks (drafting, analysis, compliance)
- Centralized orchestration and coordination mechanisms
- Communication protocols for inter-agent collaboration
- Integration interfaces for enterprise systems

4.2.2 Core Functionality

Key capabilities to be implemented include:

- **Drafting Assistance:** AI-powered contract clause generation and modification
- **Clause Classification:** Automated categorization and analysis of contract terms
- **Compliance Checks:** Regulatory compliance monitoring and validation
- **Explainability Features:** Transparent reasoning and decision justification

4.3 Evaluation Framework

A comprehensive evaluation framework will assess the effectiveness of the MAS approach across multiple dimensions.

4.3.1 Quantitative Metrics

Performance evaluation will include:

- **Task Completion Time:** Efficiency measurement for different CLM workflows
- **Accuracy Metrics:** Precision, recall, and F1-scores for classification tasks
- **Compliance Adherence:** Rate of regulatory compliance achievement
- **System Performance:** Throughput, latency, and resource utilization

4.3.2 Qualitative Metrics

User experience assessment through:

- **Explainability Evaluation:** User comprehension of AI decisions
- **Usability Studies:** Interface design and workflow efficiency
- **Trust Assessment:** User confidence in system recommendations
- **Acceptance Measurement:** Willingness to adopt MAS-based solutions

4.4 Comparative Analysis

A systematic comparison will be conducted between single-agent and multi-agent approaches under controlled scenarios.

4.4.1 Baseline Systems

Comparison will include:

- Traditional rule-based CLM systems
- Single-agent LLM-based approaches
- Hybrid systems combining rule-based and AI components

4.4.2 Evaluation Scenarios

Controlled testing scenarios will cover:

- Standard contract types across different industries
- Varying complexity levels and regulatory requirements
- Different user expertise levels and interaction patterns
- Stress testing under high-volume conditions

4.5 User Study Validation

A user study with industry professionals will validate the practical applicability and user acceptance of the MAS approach.

4.5.1 Participant Selection

Study participants will include:

- Contract managers from various industries
- Legal professionals with CLM experience
- IT professionals involved in enterprise software selection

4.5.2 Study Design

The user study will employ:

- Task-based evaluation scenarios
- Think-aloud protocols for usability assessment
- Post-task questionnaires for trust and acceptance measurement
- Comparative preference studies between different approaches

4.6 Data Collection and Analysis

4.6.1 Data Sources

Research data will be collected from:

- System performance logs and metrics
- User interaction recordings and feedback
- Expert interviews and focus groups
- Benchmark datasets for contract analysis tasks

4.6.2 Analysis Methods

Data analysis will employ:

- Statistical analysis for quantitative performance metrics
- Qualitative content analysis for user feedback
- Comparative statistical tests for baseline comparisons
- Thematic analysis for identifying patterns in user behavior

5 Expected Contributions

5.1 Scientific Contributions

This research will contribute to the academic literature through several key scientific advances:

5.1.1 Framework for MAS Design in Regulated Enterprise Workflows

The thesis will develop a comprehensive framework for designing multi-agent systems specifically tailored to enterprise workflows in regulated environments. This framework will address:

- Agent specialization strategies for compliance-sensitive tasks
- Orchestration patterns for enterprise-grade reliability and scalability
- Integration methodologies for existing enterprise infrastructure
- Design principles for maintaining audit trails and regulatory compliance

5.1.2 Comparative Empirical Results

The research will provide rigorous empirical evidence comparing single-agent versus multi-agent approaches in enterprise CLM contexts. This contribution includes:

- Performance benchmarks across multiple evaluation dimensions
- Statistical analysis of effectiveness differences
- Identification of scenarios where MAS approaches provide significant advantages
- Evidence-based guidelines for system architecture decisions

5.1.3 Evaluation Methodology for Compliance-Aware AI Systems

A novel evaluation methodology will be developed specifically for assessing AI systems in compliance-sensitive enterprise environments. This methodology will encompass:

- Metrics for measuring regulatory compliance adherence
- Frameworks for assessing explainability in multi-agent contexts
- User trust measurement techniques for enterprise AI systems
- Validation protocols for ensuring system reliability and auditability

5.2 Practical Contributions

The research will provide direct practical value to the enterprise software development community:

5.2.1 Architecture Guidance for CLM Platform Development

The findings will directly inform the architecture and design of next-generation CLM platforms by providing:

- Proven coordination strategies for multi-agent CLM workflows
- Tested explainability mechanisms for enterprise user acceptance
- Scalability patterns for handling complex contract management scenarios
- Integration best practices for enterprise system compatibility

5.2.2 Prototype Foundation for MVP Development

The developed prototype will serve as the foundation for advanced CLM platform features, offering:

- Validated multi-agent architecture patterns
- Tested user interface and interaction paradigms
- Proven performance characteristics under realistic workloads
- Demonstrated compliance and explainability capabilities

5.2.3 Industry Best Practices

The research will establish best practices for:

- Implementing MAS in enterprise software environments
- Balancing system complexity with user trust and acceptance
- Designing compliance-aware AI systems for regulated industries
- Evaluating and validating enterprise AI system effectiveness

5.3 Research Impact

5.3.1 Academic Impact

The research will contribute to multiple academic domains:

- **Multi-Agent Systems:** Novel applications in enterprise workflows
- **Enterprise AI:** Frameworks for compliance-aware system design
- **Human-Computer Interaction:** Trust and explainability in enterprise contexts
- **Legal Technology:** AI applications in contract management

5.3.2 Industry Impact

The practical contributions will benefit:

- **CLM Software Vendors:** Architecture guidance for next-generation platforms
- **Enterprise Organizations:** Improved contract management capabilities
- **Regulatory Bodies:** Frameworks for AI system compliance assessment
- **AI System Developers:** Best practices for enterprise AI implementation

5.4 Publication Strategy

The research findings will be disseminated through:

- **Conference Papers:** Submission to relevant AI and enterprise software conferences

- **Journal Articles:** Publication in multi-agent systems and enterprise AI journals
- **Industry Reports:** Practical guidance documents for enterprise software developers
- **Open Source Contributions:** Framework components and evaluation tools

5.5 Future Research Directions

This work will establish foundations for future research in:

- Advanced MAS orchestration for complex enterprise workflows
- Explainable AI techniques for multi-agent systems
- Compliance-aware AI system design methodologies
- User trust and acceptance in enterprise AI applications

6 Schedule and Work Plan

6.1 Project Timeline

The research will be conducted over a six-month period, structured into distinct phases with specific deliverables and milestones.

6.1.1 Phase 1: Literature Review and Problem Framing (Months 1-2)

Duration: 8 weeks
Objectives: Establish theoretical foundation and refine research scope
Key Activities:

- Comprehensive literature review on multi-agent systems in enterprise contexts
- Analysis of existing CLM systems and their limitations
- Review of compliance-aware AI frameworks and evaluation methodologies
- Refinement of research questions and hypotheses
- Initial system architecture design and requirements specification

Deliverables:

- Literature review report
- Refined research questions and methodology
- Initial system architecture specification
- Project proposal finalization

6.1.2 Phase 2: Methodology Development and System Design (Months 3-4)

Duration: 8 weeks
Objectives: Develop detailed methodology and system architecture
Key Activities:

- Detailed MAS architecture design for CLM workflows
- Specification of agent roles, responsibilities, and communication protocols
- Development of evaluation framework and metrics
- Design of user study protocols and experimental procedures
- Initial prototype architecture and technology stack selection

Deliverables:

- Detailed system architecture document
- Evaluation framework specification
- User study protocol design
- Technology stack and development environment setup

6.1.3 Phase 3: Implementation and Data Collection (Months 5-6)

Duration: 8 weeks
Objectives: Build prototype and conduct experimental evaluation
Key Activities:

- Prototype development and implementation
- Baseline system implementation for comparison
- Experimental data collection and performance measurement
- User study execution with industry professionals
- Initial data analysis and preliminary results

Deliverables:

- Functional MAS-based CLM prototype
- Baseline comparison systems
- Experimental datasets and performance metrics
- User study results and feedback

6.1.4 Phase 4: Analysis and Evaluation (Months 7-8)

Duration: 8 weeks
Objectives: Analyze results and validate research hypotheses
Key Activities:

- Comprehensive data analysis and statistical evaluation
- Comparative analysis between MAS and baseline approaches
- User trust and acceptance assessment
- Compliance adherence evaluation
- Results interpretation and hypothesis validation

Deliverables:

- Statistical analysis results
- Comparative evaluation report
- User acceptance and trust assessment
- Compliance effectiveness analysis

6.1.5 Phase 5: Writing and Finalization (Months 9-10)

Duration: 8 weeks
Objectives: Document findings and complete thesis
Key Activities:

- Thesis writing and documentation
- Results presentation and discussion
- Contribution analysis and future work identification
- Thesis review, revision, and finalization
- Preparation for thesis defense

Deliverables:

- Complete thesis document
- Research contribution summary
- Future work recommendations
- Thesis defense presentation

6.2 Risk Assessment and Mitigation

6.2.1 Technical Risks

Risk: Prototype development complexity exceeding timeline
Mitigation: Incremental development approach with early validation milestones
Risk: Integration challenges with enterprise systems
Mitigation: Focus on standard interfaces and protocols, maintain abstraction layers

6.2.2 Research Risks

Risk: Insufficient user study participation
Mitigation: Early engagement with industry contacts, multiple recruitment channels
Risk: Limited access to proprietary CLM data
Mitigation: Use publicly available contract datasets, synthetic data generation

6.2.3 Timeline Risks

Risk: Unexpected delays in implementation
Mitigation: Buffer time allocation, flexible milestone adjustments
Risk: Analysis complexity exceeding expectations
Mitigation: Early data exploration, simplified analysis approaches if needed

6.3 Resource Requirements

6.3.1 Computational Resources

- Development environment for MAS prototype
- Cloud computing resources for experimental evaluation
- Data storage and processing capabilities

6.3.2 Human Resources

- Access to industry professionals for user studies
- Academic supervision and guidance
- Technical consultation for enterprise system integration

6.3.3 Data and Materials

- Contract datasets for training and evaluation
- Access to existing CLM systems for baseline comparison
- Regulatory compliance documentation and guidelines

6.4 Success Metrics

Project success will be measured through:

- **Technical Milestones:** Prototype functionality and performance targets
- **Research Outcomes:** Statistical significance of comparative results
- **User Acceptance:** Positive feedback from industry professionals
- **Academic Contribution:** Novel insights and frameworks developed
- **Timeline Adherence:** Completion of phases within specified timeframes

List of Figures

List of Tables