A Novel Multi-Agent Framework for Legal Document Discrepancy Detection

Athira Raghumadhavan, Ishan Mandlik, Mandar Burande
Vijay Karanjkar, Vignesh Mohana Velu
{araghuma, imandlik, mburande, vkaranjk, vmohanav}@asu.edu

Abstract

This proposal outlines the problem statement, research objectives, methodology, expected contributions, and a structured project plan for execution. The framework leverages multiple intelligent agents to identify discrepancies in legal documents, ensuring compliance and consistency. By integrating techniques like Retrieval-Augmented Generation (RAG) and Chain-of-Thought (CoT) reasoning, the system aims to enhance accuracy and trustworthiness.

1 Introduction & Problem Statement

Legal documents should be accurate and consistent. These documents, ranging from contracts and agreements to regulatory filings, and court proceedings, often contain complex language and intricate details that must be reviewed with the utmost care to ensure compliance and avoid potential disputes. However, manually reviewing these documents is time-consuming and error-prone and requires significant knowledge and experience. This underscores the need for the detection of discrepancies in legal documents.

This multi-agentic legal document discrepancy detection framework leverages the power of Large Language Models to automate the identification of inconsistencies, errors, and omissions within legal documents while verifying compliance with federal and state laws. We cross-reference related documents and relevant laws to highlight the discrepancies that otherwise might go unnoticed.

In this project, we will explore the significance of discrepancy detection of legal documents and its applications. We discuss the benefits it offers legal professionals, including improved accuracy, reduced workload, and enhanced compliance. In addition, we will examine the challenges involved in implementing such systems. Specifically, we address the problem of detecting erroneous statements within legal documents, whether they contradict themselves within the same document or conflict with laws stated in the U.S. Code or state regulations.

2 Research Objectives & Questions

The primary goal of this project is to enhance the accuracy and efficiency of legal discrepancy detection through a multi-agent framework that can adapt to diverse legal contexts. The success of the proposed approach will be measured by its ability to improve key tasks, including legal classification, contradiction detection, and compliance verification. The multi-agent framework improves existing models by dynamically adapting and using retrieval-based knowledge augmentation and real-time agent collaboration. Its hierarchical structure enhances explainability and modular improvements, with specialized agents handling distinct tasks of the legal contract review process.

There is not much work done on the specific problem, but the existing models face generalization issues due to jurisdictional differences, static methods, and lack of interpretability, making it difficult to understand and detect discrepancies. Scalability is also a challenge when processing large, diverse legal datasets.The framework generalizes by using retrieval-augmented generation (RAG) and LLM-based reasoning. The Statute Finder ensures context-aware reasoning, and the Orchestrator Agent enables real-time collaboration across classification, compliance, and contradiction detection, making it adaptable across various legal systems.

3 Methodology

3.1 Dataset & Data Creation

The project will utilize the following existing datasets. The Atticus Project consists of three datasets: ACORD (Wang et al., 2025), a clause retrieval dataset; MAUD (Wang et al., 2023), a merger agreement dataset; and CUAD (Hendrycks et al., 2021), a commercial contract dataset. Additionally, LegalBench (Guha et al., 2023) serves as a benchmark designed to measure legal reasoning in large language models.

Furthermore, we will explore other relevant datasets that align with the objectives of this project.

3.2 Model & Approach

The proposed system uses a multi-agent architecture with transformer-based models to handle NLP tasks. It includes an Orchestrator for workflow control, a Pre-processor Agent to structure legal text, a Knowledge Agent to retrieve relevant legal information, a Compliance Checker Agent to detect contradictions and legal implications, a Clause Rewriter Agent to suggest compliant clause changes, and a Post-processor Agent to generate summaries and suggest improvements.

3.2.1 Baseline Models for Comparison

To benchmark our approach, we will compare against the following models:

- **Traditional machine learning models**: These models, such as Logistic Regression (LR) and Support Vector Machines (SVM), can be trained to classify legal documents or identify specific clauses. However, they typically require extensive feature engineering and may not capture the nuanced relationships between words and concepts in legal texts. They also suffer when it comes to adaptability and struggle in trying to transfer capabilities to domains for which they haven't been explicitly coded for.

- **Existing NLP models for legal document analysis**: Transformer models like LEGAL-BERT (Chalkidis et al., 2020) have shown promise in legal NLP tasks, including document classification and similarity detection. These models can automatically learn contextual representations of words and sentences, reducing the need for manual feature engineering. However, they may still struggle with long-range dependencies and complex reasoning required for discrepancy detection.

The limitations of these baselines include difficulties in handling complex legal language, limited generalization, and inability to adapt to evolving legal standards.

3.2.2 Proposed Model Architecture

Our approach introduces a novel multi-agentic architecture Figure 1 that leverages the strengths of specialized agents to address the challenges of legal document discrepancy detection. This includes enhanced accuracy, improved efficiency, and adaptability to different legal contexts. The agents will operate with minimal human intervention, dynamically adapting to new legal scenarios and data. (Fourney et al., 2024)

- **Orchestrator**: It creates or updates a ledger of tasks, analyzes clauses, looks up legal clauses, makes educated guesses, and creates a task plan (as shown in the attached image). We aim to use LangGraph, a tool built on top of LangChain specifically for building multi-agent systems with nonlinear complex workflows to dynamically route tasks based on context and intermediate outputs. It ensures that tasks are executed in the correct order, by the appropriate agents, and according to the evolving needs of the system, leading to better decision-making, faster processing times, and more accurate results.

- **Pre-processor Agent**: The Preprocessing Agent transforms raw legal documents into structured representations for downstream analysis. It integrates specialized agents: PDFParserAgent/OCRAgent (using Tesseract OCR) for text extraction, TextClassificationAgent (leveraging BERT, RoBERTa, or DistilBERT) for document classification and context extraction, ClauseExtractorAgent (spaCy or a parser tool) for clause extraction, and HybridNERAgent (LexNLP, Blackstone) for legal entity recognition. It interacts with the Task Planner, which assigns tasks based on document type, and the Context Bank, which stores and retrieves relevant legal precedents and contextual data.

- **Knowledge Agent**: It has a 3-layer retrieval system, including access to the Google Search Engine, a public statute finder API by LegiScan and the context bank to store relevant documents. The Agent also prioritizes the U.S. Code website for information about laws. This agent helps other agents gain contextual information about the laws and also extracts statutes and precedents for any required context.

- **Compliance Checker Agent**: It collates information from the Context Bank and the Knowledge Agent. With access to a statutory validator, precedent analyzer and contractual consistency engine for cross referencing the legal

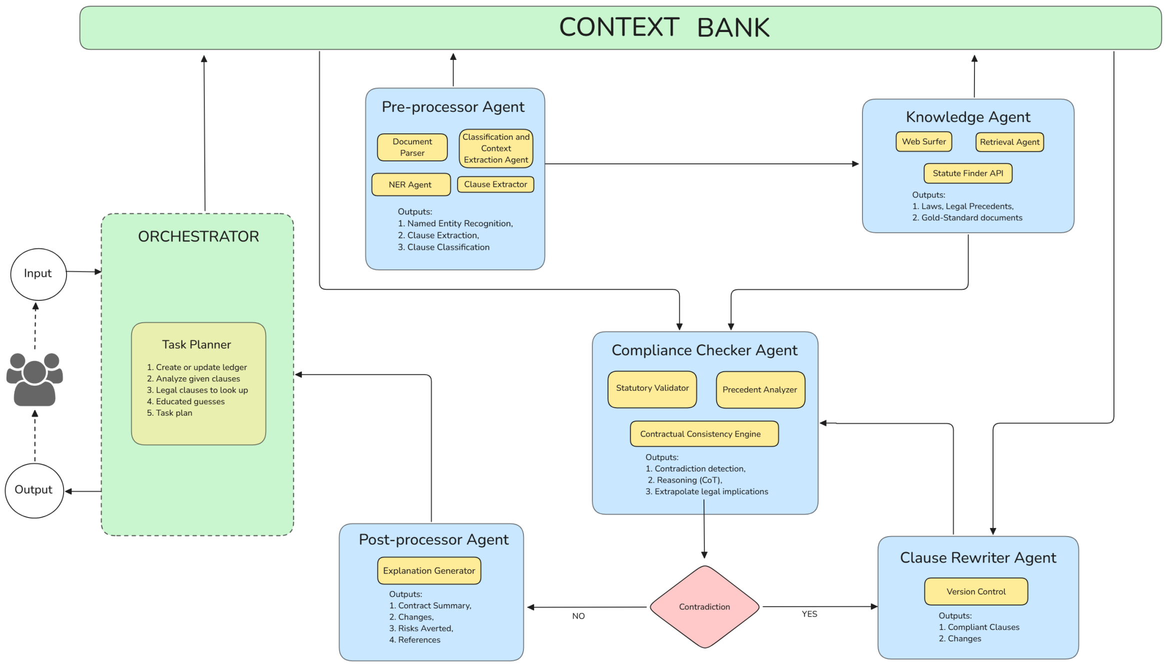

Figure 1: Proposed Architecture

document with knowledge from the Knowledge Agent, it implements a hypergraph-based contradiction detection and uses long-context LLMs for long-range dependency analysis. This agent outputs contradiction detection, reasoning using Chain-of-Thought (CoT), and extrapolation of legal implications. It flags clauses that contradict an existing statute, a set precedent or another clause from the contract and also highlights the non-compliant section of the clause and returns the law or clause being violated. The agent will also utilize reasoning capabilities to extrapolate legal implications of each violation. All of this information will be passed to the Clause Rewriter Agent to iterate over until the contradiction is comprehensively and satisfactorily resolved.

- **Clause Rewriter Agent**: It compares the input clause with pre-approved templates to ensure alignment with industry standards and best practices. It retrieves relevant template clauses using semantic search and highlights deviations or inconsistencies while preserving the clause's intent. The agent leverages the RAG Agent to enhance clause rewriting by retrieving relevant legal precedents and contextual information. Additionally, the agent flags ambiguities or conflicts within the rewritten clause and proposes edits for enhanced clarity and legal precision.

- **Post-processor Agent**: This agent creates a contract summary, lists the changes made, highlights risks averted, and provides references.

- **Context Bank**: This uses shared memory that can be accessed by all agents to help with coordination and supplement agent-to-agent data communication. It is responsible for storing the original legal document, context and classification information as well as the information retrieved from the web.

3.3 Training & Evaluation

The multi-agent system will be trained using supervised learning techniques, leveraging annotated legal documents, gold-standard contracts, and relevant case precedents to fine tune named entity recognition (NER), clause extraction, and contradiction detection tasks. The team aims to use few-shot learning and Chain-of-Thought (CoT) prompting to augment the framework's reasoning coherence when analyzing complex legal clauses.

The team plans to primarily, evaluate the model on LegalBench (Guha et al., 2023), a comprehensive evaluation benchmark consisting of different legal reasoning tasks with complex jargon, long contexts, sophisticated multi-step reasoning, intricate structure, and minimal labeled data. Legal-

Bench mainly covers three cognitive levels: 1) Legal Knowledge Memory: Testing the ability of remembering necessary legal concepts, terms, articles, and facts; 2) Legal Knowledge Understanding: assessing whether large language models can understand and interpret entities, events, and relationships in legal texts; 3) Legal Knowledge Application: evaluating the capability for correctly utilizing and reasoning with their legal knowledge to solve different legal tasks in real scenarios.

An end-to-end evaluation will comprehensively assess the framework's overall performance using a combination of accuracy, text generation metrics, and language modeling measures. Legal clause and text generation tasks will be evaluated using BLEU, ROUGE, and METEOR.

In addition to quantitative metrics, a Manual Evaluation will be conducted to review flagged contradictions, rewritten clauses, and generated summaries to assess legal accuracy, reasoning coherence, and compliance. This feedback will be incorporated to iteratively refine the framework.

4 Project Plan & Timeline

The project is structured as follows:

Tasks	Duration
Literature Review, Data Collection	2 weeks
Agent Design and Implementation	4 weeks
Framework Training & Fine-Tuning	3 weeks
Evaluation, Error Analysis	2 weeks
Refinement, Final Report	2 weeks

Table 1: Project Timeline

5 Expected Contributions & Impact

This project will introduce an innovative multi-agent architecture specifically designed for legal document discrepancy detection, automating the detection of compliance issues, inconsistencies, and contradictions. By leveraging intelligent automation, it aims to improve compliance, minimize legal risks, and streamline contract management while delivering detailed justifications and reasoning behind each detected discrepancy, thereby enhancing trust and usability for legal practitioners. Additionally, it will provide legal professionals with improved accuracy, reduced workloads, and enhanced compliance verification.

6 Related Work

The application of NLP in legal domains has led to the development of specialized models for improved text understanding. Legal-BERT (Chalkidis et al., 2020) extends BERT for legal text processing, outperforming general models like BERT (Devlin et al., 2019) in classification and contradiction detection tasks. The LawBench benchmark (Fei et al., 2023) further enhances legal reasoning evaluation across textual entailment and fact-checking.

Identifying semantic inconsistencies in legal texts remains a key challenge. Zhong et al. (Zhong et al., 2022) employ pre-trained transformers for contradiction detection, while Bommarito et al. (II and Katz, 2022) leverage contrastive learning to detect clause variations in contracts.

Recent advances introduce multi-agent systems for legal AI. Dubetcky (Dubetcky, 2024) proposes a modular framework integrating document parsing, classification, and discrepancy detection, while TrueLaw AI (AI, 2024) explores retrieval-based clause rewriting. Additionally, Wu et al. (Qin and Sun, 2024) demonstrate the effectiveness of LoRA techniques for improving contract analysis and compliance detection.

7 Potential Future Extensions

Future avenues for research include using specialized, fine-tuned models for the agents instead of the general foundational models in the current architecture. This change should lead to higher efficiency along with improved performance in nuanced and niche legal tasks.

Additionally, the model will be adapted to support multiple languages and legal systems, incorporating jurisdiction-specific modules to address variations in legal terminologies and compliance requirements. It will also incorporate multimodal capabilities to analyze legal documents containing images, tables, and diagrams by leveraging advanced OCR techniques and vision-language models for comprehensive document understanding.

References

TrueLaw AI. 2024. Agentic frameworks in legal ai systems. *AI & Law Industry Report*.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. Legal-bert: The muppets straight out of law school. *Preprint*, arXiv:2010.02559.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arXiv:1810.04805.

Alexander Dubetcky. 2024. Multi-agent legal framework for document analysis and discrepancy detection. *Artificial Intelligence & Law Journal*.

Zhiwei Fei, Xiaoyu Shen, Dawei Zhu, Fengzhe Zhou, Zhuo Han, Songyang Zhang, Kai Chen, Zongwen Shen, and Jidong Ge. 2023. Lawbench: Benchmarking legal knowledge of large language models. *Preprint*, arXiv:2309.16289.

Adam Fourney, Gagan Bansal, Hussein Mozannar, Cheng Tan, Eduardo Salinas, Erkang (Eric) Zhu, Friederike Niedtner, Grace Proebsting, Griffin Bassman, Jack Gerrits, Jacob Alber, Peter Chang, Ricky Loynd, Robert West, Victor Dibia, Ahmed Awadallah, Ece Kamar, Rafah Hosn, and Saleema Amershi. 2024. Magentic-one: A generalist multi-agent system for solving complex tasks. Technical Report MSR-TR-2024-47, Microsoft.

Neel Guha, Julian Nyarko, Daniel E. Ho, Christopher Ré, Adam Chilton, Aditya Narayana, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel N. Rockmore, Diego Zambrano, Dmitry Talisman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory M. Dickinson, Haggai Porat, Jason Hegland, Jessica Wu, Joe Nudell, Joel Niklaus, John Nay, Jonathan H. Choi, Kevin Tobia, Margaret Hagan, Megan Ma, Michael Livermore, Nikon Rasumov-Rahe, Nils Holzenberger, Noam Kolt, Peter Henderson, Sean Rehaag, Sharad Goel, Shang Gao, Spencer Williams, Sunny Gandhi, Tom Zur, Varun Iyer, and Zehua Li. 2023. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Preprint*, arXiv:2308.11462.

Dan Hendrycks, Collin Burns, Anya Chen, and Spencer Ball. 2021. Cuad: An expert-annotated nlp dataset for legal contract review. *Preprint*, arXiv:2103.06268.

Michael Bommarito II and Daniel Martin Katz. 2022. Gpt takes the bar exam. *Preprint*, arXiv:2212.14402.

Weicong Qin and Zhongxiang Sun. 2024. Exploring the nexus of large language models and legal systems: A short survey. *Preprint*, arXiv:2404.00990.

Steven H. Wang, Antoine Scardigli, Leonard Tang, Wei Chen, Dimitry Levkin, Anya Chen, Spencer Ball, Thomas Woodside, Oliver Zhang, and Dan Hendrycks. 2023. Maud: An expert-annotated legal nlp dataset for merger agreement understanding. *Preprint*, arXiv:2301.00876.

Steven H. Wang, Maksim Zubkov, Kexin Fan, Sarah Harrell, Yuyang Sun, Wei Chen, Andreas Plesner, and Roger Wattenhofer. 2025. Acord: An expert-annotated retrieval dataset for legal contract drafting. *Preprint*, arXiv:2501.06582.

Xin Zhong, Richard Hunter, and David Smith. 2022. Identifying contradictions in legal proceedings using natural language models. *Legal NLP Workshop at NeurIPS*.