

A
Project Report
ON
Real Estate Contract Intelligence



Submitted as a part of the curriculum requirements for

Master's Degree
In
Management Information System (STEM) / Business Analytics (STEM)



Submitted by

Harshal Pawar (MSBA)
Patrick Chen (MIS)
Sagar Naduvinkeri (MSBA)
Soumaya Hassan (MSBA)
Vijay Kumar Banoth (MIS)

Guided by

Prof. Dominic Sellitto
MGS-636: APPLIED AI FOR MANAGERS

SCHOOL OF MANAGEMENT
UNIVERSITY AT BUFFALO
2024-2025

Table Of Contents

Table of Contents	Error! Bookmark not defined.
1. Team Introduction	3
2. Executive Summary	4
3. Problem Statement.....	4
4. Project Objectives	5
5. Scope	5
6. System Architecture	7
Technologies Used:	7
Workflow:	7
7. Prompt Engineering Strategy.....	8
8. Example Output	10
9. Challenges and Learnings	11
10. Future Enhancements	11
11. Conclusion	12

1. Team Introduction

- **Harshal Pawar:** I'm Harshal Vijay Pawar, a Master's student in Business Analytics at the UB, graduating in June 2025. Experienced as an SAP ABAP professional for around 3 years, I lay a strong foundation in data analysis and process optimization, I'm passionate about using analytics to drive business decisions.
- **Patrick Chen:** Fresh from UB's undergraduate program, I entered the master's program for Management Information Systems. I have a passion for solving complex product issues. My expertise in client communication, problem solving, and leadership aligns with product management skills, where I plan to use to deliver impactful solutions.
- **Sagar:** Experienced analyst with a passion to solve business problems with data. My background spans database management, stakeholder collaboration and business intelligence dashboards. What makes me different? I combine technical rigor with business empathy — and I truly believe that good data can lead to better decisions.
- **Soumaya Hassan:** I'm a master's in business Analytics candidate at the University at Buffalo, with hands-on experience tackling real-world business problems as a Business Consultant Intern. In this role, I assessed client challenges and delivered strategic, data-backed solutions—honing my skills in data analysis, stakeholder communication, and cross-functional collaboration. My professional journey, including my customer-facing role at Kay Jewelers, has shaped my strengths in customer relationship management, while fueling my interest in the practical applications of AI/ML in business strategy.
- **Vijay Kumar Banoth:** I am an ITIL-certified tech enthusiast with nearly 6 years of experience, have a passion for solving complex product issues. My expertise in Azure Cloud and configuration lifecycle management, a subset of Product Lifecycle Management, positions me to transition into product management, where I can use my problem-solving skills and database ability to transition into product management and deliver impactful solutions.

2. Executive Summary

The **AI Contract Analyzer for Real Estate agreements** is designed to reduce the cost, complexity, and inefficiencies associated with reviewing real estate agreements. Built using the Mistral &B language model through Ollama and integrated with Python & PyMuPDF, the solution enables instant analysis of rental agreements by extracting, summarizing, and flagging essential clauses. Operating entirely offline, it ensures data privacy while delivering rapid, accurate insights into lease terms, obligations, and potential risks. This project targets the underutilized domain of AI in legal document analysis, specifically focused on lease and property contracts.

3. Problem Statement

Small businesses and private property owners typically have a disproportionately high expense in terms of legal costs associated with lease agreements. While new business applications in the U.S. were at all-time highs, over **5.5 million in 2022**, entrepreneurs are entering the market with little initial capital, averaging just **\$10,000**. But lease review legal costs can approach over \$13,000, more than some spend launching their whole business.

This economic cost is aggravated by three core problems:

Complex Lease Structures: Lease agreements are verbose and filled with technical legal jargon, making them incomprehensible to non-experts. One requires legal expertise to understand the full meaning of expressions like escalation clauses, use restrictions, or indemnity provisions.

Hidden or Baked-in Terms: Essential aspects such as penalties, utility charges, or baked-in fees tend to be concealed in lengthy legal terminology. Without professional advice or considerable time spent, essential provisions can be easily missed — leading to expensive consequences later.

Time-Consuming Manual Review Process: Manual review of lease contracts is labor-intensive, error-prone, and expensive. The process tends to involve several rounds of interactions with legal professionals, adding to delays and driving up legal costs.

While demand for contract transparency is growing, solutions in legal tech today are mainly focused on contract drafting. Only 15% of legal tech adoptions today are aimed at contract analysis, the ACEDS survey asserts. This is an enormous need gap for understanding-oriented tools rather than simple agreement creation.

There is an urgent need for a simplified, accurate, and automated device that streamlines lease contract analysis, reducing cost, improving clarity, and minimizing risk, especially for those who are not able to pay for exclusive legal services.

4. Project Objectives

The primary goals of the AI Contract Analyzer project are:

- a. **Automated Clause Extraction:** Develop a system that identifies and extracts key clauses from real estate agreements, including lease terms, rent amounts, security deposits, utility responsibilities, and penalty clauses.
- b. **Summarization of Legal Documents:** Provide concise, human-readable summaries of complex lease agreements to facilitate better understanding for non-experts.
- c. **Risk Identification:** Implement mechanisms to detect and flag potential risks and obligations within contracts, such as hidden fees or restrictive clauses.
- d. **Offline Functionality:** Ensure the tool operates entirely offline to maintain user privacy and data security.
- e. **Prompt Engineering Utilization:** Employ prompt engineering techniques to guide the language model's output without the need for extensive retraining or fine-tuning.

5. Scope

The AI Contract Analyzer supports the following key features:

➤ **Lease term extraction (duration, monthly rent, deposits)**

- Monthly Lease Rate

Defines your recurring cost.

Example: If a lease states \$2,000/month but you mistakenly read it as \$1,200, your annual budgeting could be off by nearly \$10,000.

- Lease Term

Shows how long you're legally tied to the lease.

Example: A 3-year lease locks you in long-term, while a 6-month lease offers flexibility. This matters if your business is still scaling or may relocate soon.

- Security Deposit

Reveals upfront capital needed and refund conditions.

Example: A deposit of \$5,000 might strain a startup's initial cash flow, especially if it's not clearly refundable or tied to maintenance conditions.

➤ **Utility and maintenance responsibilities**

Avoids surprise costs.

Example: If the tenant is responsible for all utilities (electricity, water, trash), that might add \$400–\$700/month, significantly increasing actual costs beyond rent.

➤ **Penalty clauses**

Identifies hidden financial risks.

Example: A late rent fee of \$100/day could quickly escalate if you're late by even a few days—something often overlooked during review.

➤ **Property use restrictions**

Ensures legal compliance and business suitability.

Example: If the lease says "residential use only," but you plan to run a small café or office, you may violate zoning laws or lease terms.

➤ **Summary generation of obligations and liabilities**

Lease agreements often bury critical responsibilities in dense legal jargon.

Summarizing them in plain English allows users to **quickly understand their commitments and potential risks**.

It ensures users don't miss clauses that could result in legal trouble or unexpected expenses.

Example: Tenant handles interior maintenance (plumbing, electrical) and must carry \$500K liability insurance. Landlord maintains exterior and roof. Late rent incurs \$75/day fee. Early termination forfeits full deposit. Landlord not liable for tenant-related damages.

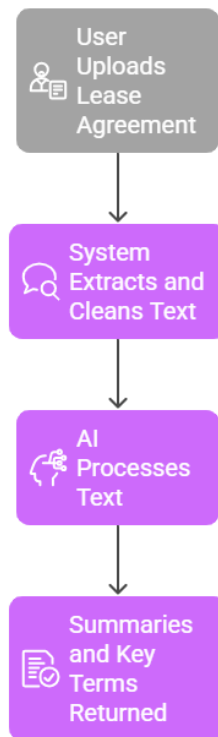
6. System Architecture

Technologies Used:

- **Ollama:** Lightweight LLM runner for local inference.
- **Mistral 7B:** Open-source language model used for contract understanding.
- **Python:** Backend application logic and integration.
- **PyMuPDF:** Document parsing and text extraction from PDFs.

Workflow:

AI-Powered Lease Agreement Processing



1. User Uploads Lease Agreement

- The user begins by uploading a lease or rental contract in digital format (PDF, DOCX, etc.).

2. System Extracts and Cleans Text

- A backend system (using Python scripts and OCR/NLP tool **PyMuPDF**) extracts raw text from the uploaded document.
- This step removes formatting issues, headers, footers, and irrelevant content to prepare clean, structured input for the AI.

3. AI Processes Text

- The cleaned text is sent to the **Mistral AI model**, accessed locally via **Ollama**.
- Engineered prompts guide the model to focus on key clauses like rent amount, termination conditions, renewal policies, deposit terms, etc.

4. Summaries and Key Terms Returned

The AI generates:

- A concise summary of the contract.
- A structured list of key terms and potential red flags.
- These results are returned to the user through the interface, aiding quick review and decision-making.

7. Prompt Engineering Strategy

Instead of retraining the LLM, our system leverages natural language prompts to guide the Mistral model in extracting structured data from lease agreements.

Prompts were structured as natural language instructions, starting with general context (e.g. “You are a legal expert”), followed by a checklist of specific extraction tasks.

Key characteristics of the prompt engineering strategy include:

- **Clearly scoped prompts outlining extraction tasks:** Each prompt lists the information to extract.
- **Defined parameters for lease terms and legal attributes:** domain-specific parameters based on standard legal clauses.
- **Prompt phrasing** emphasizes locating specific phrases such as ‘*tenant shall be responsible for...*’ or ‘*a fee of \$x.xx shall apply if...*’ to deal with compound clauses.
- **Sequential parsing** to reduce ambiguity and reinforce consistency. Prompts were broken into sub-tasks to reduce noise and ensure consistent output.

Example: Extract key fields like "Monthly Rent", "Lease Duration", "Security Deposit", etc., along with their respective values.

```
# --- Prompt Template Function ---
def generate_prompt(contract_text):
    return f"""
You are a legal expert.

Please extract the following from the lease agreement:


1. Monthly Lease Rate
2. Length of Lease Term
3. Security Deposit Amount
4. Utilities Responsibility
5. Penalties/Fees
6. Property Use

Return it in this format:
- Monthly Lease Rate:
- Lease Term:
- Security Deposit:
- Utilities:
- Penalties/Fees:
- Property Use:

Contract:
{contract_text}
"""
```

This template ensures that the outputs can be easily parsed or displayed. It is designed to analyze different clause structures commonly found in lease documents.


8. Example Output



AI Contract Analyzer


Upload a lease agreement and extract key legal terms using Mistral via Ollama.

Upload PDF



Drag and drop file here
Limit 200MB per file • PDF


Browse files



AI Contract Analyzer


Upload a lease agreement and extract key legal terms using Mistral via Ollama.

Upload PDF




Drag and drop file here
Limit 200MB per file • PDF

Browse files



Rental_Lease Agreement1.pdf 422.4KB

×



Extracted Lease Terms

This document appears to be a residential lease agreement between a tenant named Jo

- Duration of the lease (unspecified)
- Monetary amounts, such as rent, security deposit, additional fees, and penalties
- Obligations and responsibilities of both the tenant and landlord regarding maint
- Rules and regulations governing the tenant's behavior and use of the property
- Addresses for contacting either party related to the tenancy
- Waivers, waivers of waiver, and indemnification clauses
- Entire agreement clause stating that this lease constitutes the entire agreement
- Signatures and acknowledgments of both the tenant and landlord, as well as a sta
- A provision allowing for the display of 'For Sale' or 'For Rent' signs on the pr
- Time being of the essence in this lease

It is important to consult with a legal professional when signing any legal documen

Field	Extracted Value
Monthly Lease Rate	\$900
Lease Term	1 year
Security Deposit	\$700
Utilities	Tenant pays all utilities
Penalties/Fees	\$15/day late fee
Property Use	Residential only

9. Challenges and Learnings

- **Text Extraction Variability:** PDF layouts varied significantly; preprocessing was crucial.
- **Ambiguity in Language:** Legal text often includes vague or compound clauses.
- **Prompt Tuning:** Iterative prompt refinement improved output reliability.
- **Resource Efficiency:** Mistral offered a good trade-off between performance and hardware cost.

Key Learnings:

- Effective prompt engineering is important in guiding language models to produce desired outputs, especially when fine-tuning is not feasible.
- Preprocessing of input documents significantly enhances the accuracy of information extraction.
- User feedback is vital in refining system outputs to meet real-world needs and expectations.

10. Future Enhancements

- **Visual Highlighting:** Highlight key clauses directly in the PDF.
- **Multi-Contract Comparison:** Enable comparison between different agreements.
- **Contract Type Expansion:** Support employment and service contracts.

- **Advanced Models:** Evaluate models fine-tuned on legal datasets like CUAD

11. Conclusion

The AI Contract Analyzer demonstrates how open-source LLMs can be harnessed to streamline legal document review in real estate. In the current landscape, where time and accuracy are important, our solution offers immediate value by streamlining the review process, reducing human error, and ensuring that vital contractual elements are not overlooked. Its offline functionality further ensures data privacy, making it a reliable choice for users concerned about confidentiality.

Looking ahead, the foundational architecture of this analyzer presents numerous opportunities for expansion. By incorporating advanced features such as visual clause highlighting, multi-contract comparison, and support for a broader range of legal documents, we can enhance its utility and applicability across various sectors. Moreover, integrating fine-tuned models on specialized legal datasets will further refine its accuracy and relevance.

In conclusion, this project not only demonstrates the practical application of AI in legal document analysis but also lays the groundwork for future innovations that can revolutionize how we interact with and interpret complex legal texts. By bridging the gap between advanced technology and everyday legal needs, the AI Contract Analyzer positions itself as a pivotal tool in the evolution of legal tech solutions.

12. Appendix

Appendix A: Presentation Slides

- **Title:** *AI Contract Analyzer – Project Presentation*
- **Description:** A slide deck summarizing the project's key aspects, including problem statements, solution approaches, technical implementations, and anticipated impacts.

Appendix B: Source Code Repository

- **Title:** *AI Contract Analyzer – Source Code*
- **Description:** The complete source code developed for the project, encompassing modules for document parsing, AI model integration, and user interface components.

Appendix C: Demonstration Video

- **Title:** *AI Contract Analyzer – Demonstration*
- **Description:** A video walkthrough showcasing the application's functionalities, from uploading a lease agreement to generating summarized outputs.

Appendix D: References

- Mistral AI. (2025). Mistral 7B [Open-source language model]. Retrieved from <https://mistral.ai>
- Ollama. (2025). Ollama: Run large language models locally. Retrieved from <https://ollama.com>
- Hendrycks, D., Burns, C., Kadavath, S., et al. (2021). CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. arXiv preprint arXiv:2103.06268. Retrieved from <https://arxiv.org/abs/2103.06268>
- OpenAI. (2025). ChatGPT (May 6 version) [Large language model]. Retrieved from <https://chat.openai.com/chat>