

Pharmacovigilance Risk Assessment using Generative AI



Introduction

Reading and comprehending contracts after their execution is a crucial step in assessing the risks associated with contracts of all natures and types. Legal experts analyze the clauses in executed contracts to identify hidden risks that could emerge or potential breaches that require remedial measures. A typical contract contains information about the parties involved, their responsibilities, rights, obligations, and various clauses such as recitals, definitions, scope of services, terms and terminations, payments, intellectual property rights, and warranties.

In the pharmaceutical industry, business contracts also include clauses related to regulatory compliance, drug safety, and legal risks. Legal experts are involved in this manually intensive process of comprehending and identifying risks present in business contracts. Based on these risks, experts may recommend a separate agreement that spells out the responsibilities of the parties involved. This process of identifying risks can be time-consuming, fraught with ambiguity, error-prone, and inconsistent.

Our solution, powered by Generative AI technology, addresses this problem by analyzing a vast repository of contracts. The solution processes contracts of different natures, extracts key clauses, entities, and obligations using advanced prompts and machine learning, provides a risk assessment score, and helps in making key decisions for the legal experts. This reduces the total effort and improves compliance. The solution also provides reasonable explanations for the inferences made by the generative AI model. In this case study, we present the solution implemented for a large pharmaceutical company, along with the results and benefits in the following sections.

Background

Navikenz's client, among the global pharmaceutical leaders, engages with different partners to market and distribute their pharmaceutical products to healthcare providers, pharmacies, and patients. Drug safety is an extremely important subject for the pharma company across business functions such as clinical trials, sales and marketing, and regulatory and compliance. In particular, the legal affairs team dealing with regulatory compliance is saddled with the responsibility of handling agreements that could have potential risks. It is therefore imperative for this legal team to clearly understand the nuances of the contracts and the responsibilities of all partners. Specifically, they should understand the partner's responsibilities, rights, and obligations regarding collecting, monitoring, alerting, and communicating adverse events information to regulatory authorities and other stakeholders to ensure compliance.

Our client's Pharmacovigilance legal team reads and understands these nuances to assess risks present in the contracts and decide whether to create a separate pharmacovigilance agreement. The legal team uses a document solution based on a predefined set of questions. The answers provided are based on keywords/phrases and have a high degree of ambiguity.

Their key decisions are based on a set of binary questions and a scorecard model that provides a risk score. The risk assessment methodology is not sufficiently conclusive as it is dependent on the user typing in the questions and their understanding of the answers.

Furthermore, our client was at a crucial juncture of exploring and adopting Generative AI in their suite of cognitive offerings. The rapidly changing AI landscape presented them with both opportunities and challenges. Our goal was to address not only the challenges and pain points of the legal team but also to create opportunities to introduce new capabilities to their existing AI platform and enhance its utility to the organization.

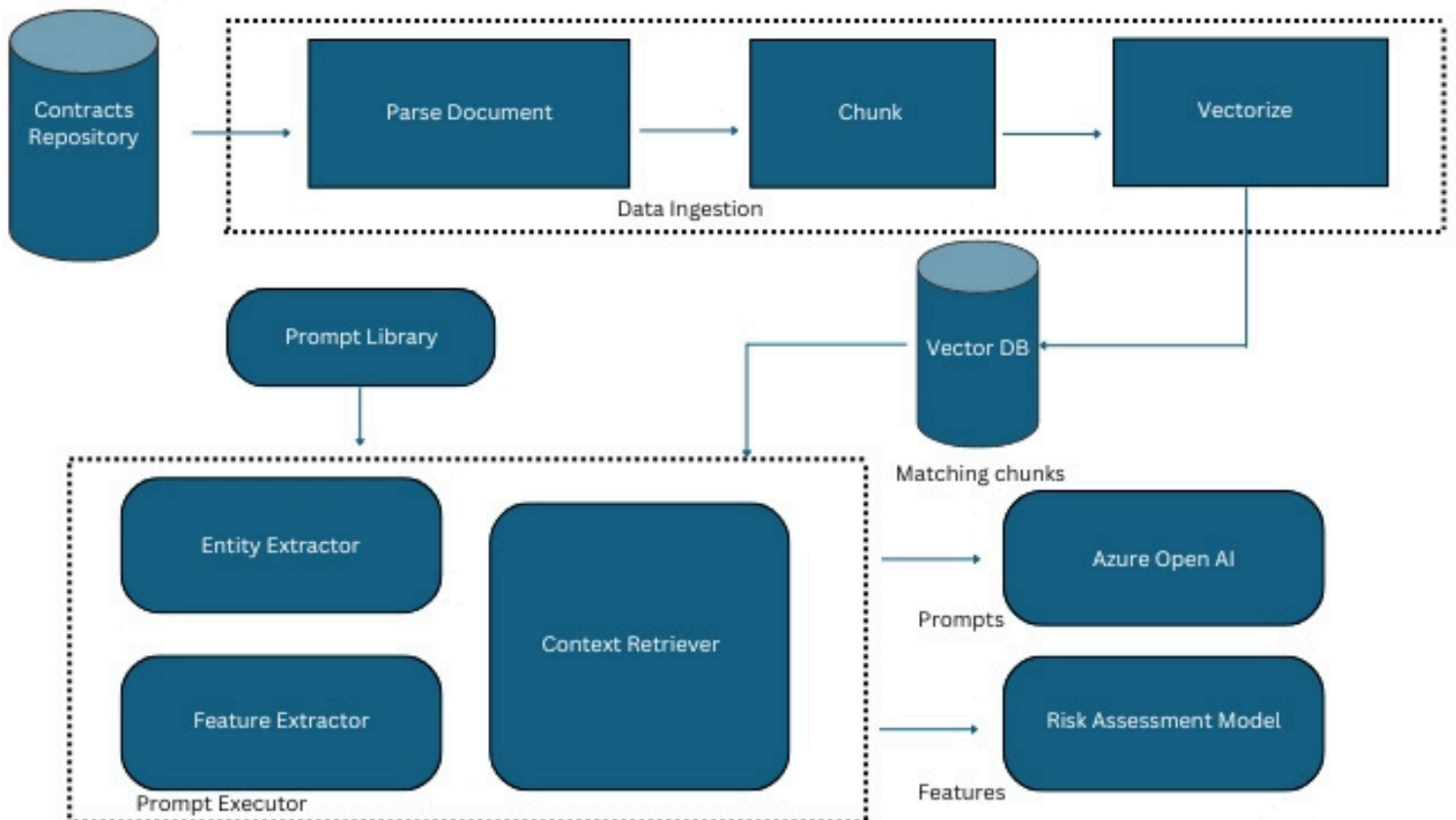
Problem

The pharma company's legal team deals with complex business agreements written in nuanced legal language in both English and non-English languages. The problem scope was confined to distribution agreements written in English, with the objective of identifying and extracting key information and clauses, and answering key questions around responsibilities, obligations, risks, and rights of the parties/partners in different territories.

Solution

The solution was conceptualized and designed to analyze and process distribution agreements while providing the capability to handle other contract types. The overall approach combined Generative AI and Machine Learning models for classifying contracts based on risks. Retrieval Augmented Generation (RAG) was at the core of the solution for various Natural Language Processing tasks. The key steps were Data Collection, Data Analysis, Prompt Engineering, Model Building, and Model Evaluation.

High Level Solution Architecture



The solution had the following components.

- Data Ingestion Pipeline :
 - 1.The contracts were parsed using pdf parser. The parsed contents were chunked and vectorized before storing them in Vector DB.
 - 2.Different chunking strategies were experimented and implemented.
- Prompt Library :
 1. A library was designed to store prompts that were finalized through experimentation. A sample prompt is shown below.
 2. It was used to manage different prompts designed for various tasks.
- Prompt Executor:
 - 1.This component was designed to construct prompts for various task using the prompt templates and relevant data fetched from Vector DB.
- Azure Open AI
 - 1.Azure Open AI was used as Large language model for various tasks.
 - 2.The prompts constructed were executed on this model.
- Risk Assessment Model
 - 1.Risk assessment model is a classification model to provide risk score indicating if Pharmacovigilance Agreement is needed. It was built using Scikit and Tensorflow.

Sample Prompt

```
Role : You are a Legal expert . You currently hold the Marketing Authorization (MAH) Rights.  
Instruction: Read the contract given below and check if there is transfer of Marketing Authorization  
Rights to other parties and answer the question.  
Rules : Determine the partner of company ABC.  
        If partner is responsible for registering the product with regulatory body or  
        If ABC is transferring the marketing authorization holder rights to partner in the distribution  
territory ,  
        then consider partner as Marketing Authorization holder  
        Else ABC is Marketing Authorization holder  
Context: {context}  
Question: Please use the rules strictly to determine the owner of Marketing Authorization Rights.  
Format : {{"MAH_Owner" : <MAH Holder>}}
```

Results & Findings

The core of the solution was built using Generative AI, making it essential to identify the right metrics and evaluate them. Most tasks were either Extraction or Classification. In the actual implementation of the project, we used precision, recall, and f1-score along with Accuracy. For simplicity, we are presenting only Accuracy.

Consistency was another important factor in evaluating the results. While most tasks were well-defined tasks such as classification and extraction of entities, tasks such as extracting a few clauses needed the consistency dimension since the produced content had varying lengths. The risk assessment model was built using Random Forest, Gradient Boosting, and Artificial Neural Network. In most cases, the model had a precision of around 90% and a recall of 95%.

Task	Type of Task	Accuracy
Type of License (In or Out)	Classify	65.00%
Marketing Authorization Holder	Assume, Look for Evidence, Decide	85.00%
Product Packaging	Look for Evidence, Decision	79.00%
Point of Contact for Local Authorities	Look for Evidence to determine who is responsible for dealing with Local authorities	63.00%
Parties	Extraction	82.00%
Territory	Extraction	81.00%
Effective date	Extraction	80.00%
Type of Contract Commercial or Service	Classify	90.00%
Risk Assessment Model	Classification	92.00%

Benefits

Our solution was able to reduce the manual effort of reading and comprehending contracts and reduce review and revision cycle time for the legal team. The document triaging helped legal experts prioritize the contracts that required more attention.

Effort Savings

Potential reduction of at least 20% in total effort of the legal team in processing business contracts.

Improved Regulatory Compliance

Ability to proactively identify obligations and mitigate the risk of missing fulfilling the obligations. Improves the organization's business risk posture, reduces financial risks, and improves compliance.

Conclusion

The solution was highly effective in solving the problem of comprehending and assessing risks in contracts. The Retrieval Augmented Generation method of using large language models and Vector DB was effective in solving NLP tasks without the need for explicit fine-tuning of LLMs or a huge corpus of contract data. Tasks related to extraction had high-quality results, while tasks related to answering key questions had good, acceptable results. Few tasks such as extracting the recitals from the contract, had some inconsistencies. In most cases, the results were impressive despite slight inconsistencies in the vector embeddings of the text.

The risk assessment model produced impressive results using machine learning approaches. However, the few-shot classification was not effective for this task, which can be attributed to the large language model's inability to understand that complex task.

In the future, the solution has the potential to address contracts of different natures with minor customization in prompts, without the need for a huge corpus for fine-tuning the large language model.