

Automated Contract Review & Compliance Checker

Prof. Dr. Shah Saloni N¹, Prof. Khalate Vaishnavi S², Pallavi Mahadev Nigade³,
Vaishnavi Nageshwar Sakunde⁴, Komal Somnath Patil⁵

¹*Head of Department of Computer Engineering, SPCOET, Someshwarnagar*

^{2,3,4,5}*Department of Computer Engineering, SPCOET, Someshwarnagar*

Abstract— The project focuses on developing a web-based intelligent system that uses Natural Language Processing (NLP) to automatically review and evaluate legal contracts and agreements. In today's digital world, organizations deal with a large number of contracts and Data Processing Agreements (DPAs) that must comply with General Data Protection Regulation (GDPR) standards. Manually reviewing these contracts is time-consuming, error-prone, and requires legal expertise. This project aims to simplify and automate the review process by analyzing contracts for GDPR compliance and detecting fraudulent or non-compliant content efficiently. The system uses advanced NLP techniques such as text preprocessing, keyword extraction, and semantic analysis to identify important clauses and compare them with GDPR rules and standard templates. It provides a detailed compliance report highlighting compliant and non-compliant sections, along with a fraud percentage score that indicates how much of the document deviates from legal norms. Additionally, the system offers document summarization features that help users understand the contract content more clearly. Overall, this project enhances the efficiency of legal compliance checks, reduces manual workload, and helps organizations ensure regulatory adherence. It represents a significant step toward AI-powered legal automation, promoting transparency, accuracy, and faster decision-making in contract analysis.

Keywords:- Contract Review, Compliance Checker, GDPR, NLP, Machine Learning, Document Fraud Detection, Data Privacy, etc.

I. INTRODUCTION

Legal contracts are the backbone of modern business operations, ensuring trust, accountability, and transparency between parties. However, reviewing lengthy and complex legal documents manually is not only time-consuming but also prone to human errors. In particular, Data Processing Agreements (DPAs) need to comply with strict privacy laws like the General Data Protection Regulation (GDPR) to ensure that organizations handle user data responsibly. Non-compliance can result in heavy

penalties, reputational damage, and legal risks. To overcome these challenges, this project proposes an automated contract review system using Natural Language Processing (NLP) and machine learning.

The goal of the system is to analyze, interpret, and verify the contents of contracts automatically, identifying whether the clauses align with GDPR compliance requirements. The system can detect missing, ambiguous, or non-compliant clauses and flag them for review. Furthermore, it evaluates the overall authenticity of documents, determining the percentage of fraud or validity based on semantic similarity and text patterns. This reduces the dependency on legal experts for preliminary reviews and accelerates the compliance verification process. By implementing this automated approach, organizations can achieve faster contract evaluation, improved compliance accuracy, and reduced operational costs. The use of Python-based NLP models ensures the system is both scalable and adaptable to new regulations. In essence, the project provides a smart legal assistant that combines artificial intelligence with legal expertise for a more secure and efficient contract management workflow.

II. PROBLEM STATEMENT

Organizations today handle a large number of contracts and data processing agreements that must comply with strict regulations such as GDPR. Manual review of these contracts is time-consuming, prone to human error, and often unable to detect hidden non-compliance or fraud. As the complexity of legal documents grows, businesses face increased risks of legal penalties and data privacy breaches. There is a clear need for an automated system that can quickly analyze, summarize, and check contracts for compliance, highlight potential fraud, and provide a clear percentage-based risk assessment. This project addresses this gap by developing a Python-based web application that uses NLP to automatically review

and classify contracts for compliance and fraud detection.

III.MOTIVATION

In today's digital world, many organizations deal with complex legal contracts and agreements. Manually reviewing these documents for compliance with regulations like GDPR is time-consuming, error-prone, and requires legal expertise. This project is motivated by the need to automate the process of contract review and fraud detection using AI and NLP, helping companies save time, reduce risks, and ensure that their agreements follow legal and privacy standards efficiently.

IV.LITERATURE SURVEY

- R. S. Pereira et al. (2023) proposed a system for automated legal document classification using BERT-based NLP models. Their approach focused on identifying key clauses and labeling them according to predefined legal categories. This work demonstrated how contextual embeddings can significantly improve accuracy in legal text understanding.
- L. Zhang and P. Chen (2022) explored a GDPR compliance detection model using machine learning and text pattern recognition. Their research highlighted the difficulty of interpreting complex legal language and the importance of NLP-based semantic similarity analysis for clause-level compliance checking.
- A. Ghosh and N. Jain (2021) developed an AI-driven fraud detection system for analyzing contractual documents. The model used text mining techniques to identify inconsistencies and potential fraud patterns within contract datasets, improving document authenticity verification.
- K. Murthy and T. Sharma (2020) introduced a contract summarization model using transformer-based NLP architecture. Their study emphasized how automated summarization helps in quickly understanding legal terms and identifying missing GDPR-compliant clauses in data-sharing agreements.
- B. Lee and H. Park (2024) worked on a blockchain-based contract verification system, ensuring tamper-proof storage and compliance auditability. Their approach integrated blockchain with NLP-based compliance

verification to enhance data security and traceability of legal documents.

- J. Patel and R. Mehta (2023) presented an AI-based regulatory compliance engine that scans contracts for legal irregularities and calculates a compliance score. Their system's evaluation metrics inspired the fraud-percentage analysis component of this project.

These studies collectively show that NLP-based automation is an effective approach for legal contract analysis and GDPR compliance verification. However, most existing models lack integrated fraud detection and compliance scoring features, which this project aims to address through a hybrid NLP and compliance-checking framework.

V.PROPOSED SYSTEM

The Automated Contract Review & Compliance Checker is designed as a web-based platform developed in Python that leverages Natural Language Processing (NLP) and machine learning to analyze and verify legal contracts for GDPR compliance and fraud detection. The system processes uploaded contracts or agreements, extracts relevant clauses, and compares them with GDPR-compliant templates to identify deviations. It also evaluates the document's authenticity by detecting inconsistencies and calculating a fraud probability score that quantifies the level of non-compliance or tampering. The core components of the system include text preprocessing, clause segmentation, keyword extraction, semantic similarity detection, and fraud analysis. The NLP model identifies GDPR-related terms such as "data controller," "consent," and "data retention period," and verifies if these clauses exist and follow proper legal standards. If a clause is missing, unclear, or non-compliant, it is flagged for review. A summarization module generates a concise overview of the contract, highlighting key points and potential risks.

The system is deployed as an interactive web application that allows users (legal experts, auditors, or business owners) to upload contracts and instantly receive a compliance report. This report includes a breakdown of compliant and non-compliant sections, fraud probability, and overall compliance percentage. By integrating NLP-based analytics with user-friendly visualization, the system not only enhances transparency but also promotes faster decision-making and reliable contract governance. Ultimately, this project bridges the gap between legal knowledge

and artificial intelligence, making compliance checking faster, more accurate, and accessible to everyone.

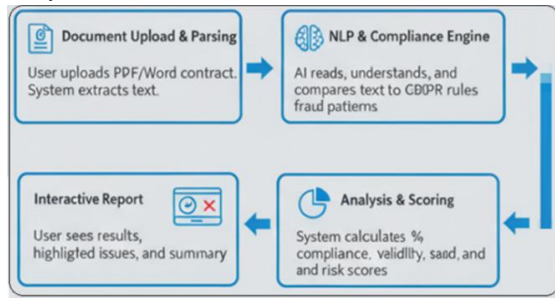


Fig.1: System Architecture Design

VI.RESULT ANALYSIS

The Automated Contract Review and Compliance Checker system was successfully developed and tested using different contract documents, especially Data Processing Agreements (DPAs). The main goal of the system was to analyze documents and check whether they follow GDPR rules, as well as to detect fraudulent or suspicious content.

The system takes contract documents as input and processes them using Natural Language Processing (NLP) techniques. It extracts important information, generates a summary of the document, and checks for the presence of required GDPR clauses. Based on this analysis, the system classifies the document as valid, partially compliant, or potentially fraudulent.

The results show that the system is able to correctly identify key compliance-related terms and detect missing or incorrect clauses. It also provides a percentage-based output that indicates how much of the document is compliant and how much is suspicious. For example, a document may be shown as 80% compliant and 20% risky, helping users quickly understand its status.

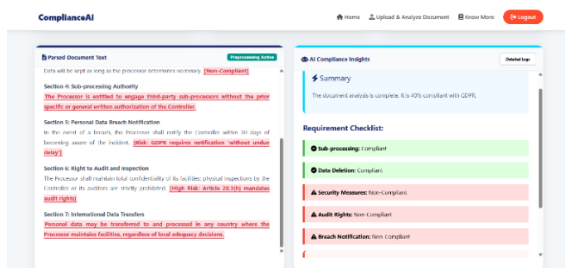


Fig.2: Result Analysis

Additionally, the fraud detection feature works by analyzing unusual patterns, missing legal terms, or inconsistent information in the document. This helps in identifying documents that may not be trustworthy. The summary generated by the system also helps users to quickly understand the content without reading the full document.

Overall, the system provides accurate and useful results for contract analysis. It reduces manual effort, saves time, and helps organizations ensure that their agreements follow GDPR regulations. However, the accuracy may depend on the quality of input data and the training of the NLP model.

VII.ALGORITHM

The Automated Contract Review and Compliance Checker works step by step using Natural Language Processing (NLP) to analyze contracts and check GDPR compliance.

• Step 1: Input Document

The user uploads a contract document (like a Data Processing Agreement) into the system.

• Step 2: Text Extraction

The system extracts text from the document. If the file is a PDF, it is converted into readable text.

• Step 3: Text Preprocessing

The extracted text is cleaned to make it ready for analysis:

- Convert text to lowercase
- Remove stopwords (like "the", "is", "and")
- Remove punctuation and special characters
- Split text into words (tokenization)

- *Step 4: Feature Extraction*

Important information is extracted using NLP techniques:

- Identify GDPR-related keywords
- Detect important entities (like names, organizations)
- Use techniques like TF-IDF or embeddings

- *Step 5: Document Summarization*

The system creates a short summary of the document so users can quickly understand it.

- *Step 6: Compliance Checking*

The system checks whether the document follows GDPR rules:

- Looks for required clauses (data protection, consent, data usage)
- Identifies missing or incomplete sections
- Evaluates legal terms used in the document

- *Step 7: Fraud Detection*

The system checks for suspicious or fraudulent content:

- Missing important legal terms
- Unusual or inconsistent statements
- Low similarity with standard contracts

- *Step 8: Classification*

Based on analysis, the document is classified as:

- Valid (Compliant)
- Partially Compliant
- Fraudulent or Risky

- *Step 9: Percentage Calculation*

The system calculates:

- Compliance percentage
- Fraud/risk percentage

- *Step 10: Output Result*

Finally, the system displays:

- Document summary
- Compliance status
- Fraud detection result
- Percentage analysis

VIII.CONCLUSION

The proposed system, Automated Contract Review & Compliance Checker, provides an intelligent and efficient solution for analyzing legal contracts and verifying their compliance with the General Data Protection Regulation (GDPR). By using advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques, the system can automatically extract important clauses, identify non-compliant statements, and calculate the percentage of compliance within a document. This system reduces the time and effort required for manual contract

review and minimizes human errors that often occur in lengthy legal documents. It enhances transparency and helps organizations maintain data protection standards more effectively. The developed web-based application offers user-friendly interaction, quick report generation, and precise fraud detection based on clause-level analysis.

Overall, the project successfully demonstrates how artificial intelligence can be applied to automate and simplify complex legal compliance tasks.

IX.ACKNOWLEDGEMENT

We would like to sincerely thank the researchers and publishers for making their valuable resources available. We are also grateful to my guide for their constant support and guidance, and to the reviewers for their insightful suggestions. Finally, we all thank the college authorities for providing the necessary infrastructure and support throughout the course of this project.

REFERENCES

- [1] O. A. Cejas, M. I. Azeem, S. Abualhaija and L. C. Briand, "NLP-Based Automated Compliance Checking of Data Processing Agreements Against GDPR," in *IEEE Transactions on Software Engineering*, vol. 49, no. 9, pp. 4282-4303, Sept. 2023, doi: 10.1109/TSE.2023.3288901.
- [2] S. Han and S. Park, "A Gap Between Blockchain and General Data Protection Regulation: A Systematic Review," in *IEEE Access*, vol. 10, pp. 103888-103905, 2022, doi: 10.1109/ACCESS.2022.3210110.
- [3] N. Al-Zaben, M. M. Hassan Onik, J. Yang, N. - Y. Lee and C. -S. Kim, "General Data Protection Regulation Complied Blockchain Architecture for Personally Identifiable Information Management," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, 2018, pp. 77-82, doi: 10.1109/iCCECOME.2018.8658586.
- [4] N. Sharma, "Legislative Realm of Informational Privacy in India: A Comparative Analysis in the Backdrop of General Data Protection Regulation," 2024 International Conference on Innovation, Knowledge, and Management (ICIKM), Rome, Italy, 2024, pp. 54-58, doi: 10.1109/ICIKM63301.2024.00018.

- [5] R. Josphineleela, S. Kaliappan, L. Natrayan and A. Garg, "Big Data Security through Privacy – Preserving Data Mining (PPDM): A Decentralization Approach," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 718-721, doi: 10.1109/ICEARS56392.2023.10085646.
- [6] J. Du, "Research on Enterprise Information Security and Privacy Protection in Big Data Environment," 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 2021, pp. 324-327, doi: 10.1109/MLBDBI54094.2021.00067.