



International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)



ijerst.editor@gmail.com
editor@ijerst.com

Research Paper

LEGAL CONTRACT ANALYZER - CLAUSE / RISK HIGHLIGHTER

¹ M Swathi, ² C Nikita, ³ D Radhika, ⁴ A Akshitha, ⁵ N Akash

¹Assistant Professor, ²³⁴⁵Students

Department of CSE(AIML)

Siddhartha Institute of Technology & Sciences, Narapally

mswathi.csm@siddhartha.org.in, 23TQ1A6632@siddhartha.co.in, 23TQ1A6643@siddhartha.co.in,
23TQ1A6659@siddhartha.co.in, 23TQ1A6641@siddhartha.co.in

ABSTRACT

The Legal Contract Analyzer — Clause/Risk Highlighter is an AI-powered document analysis system designed to help users understand and review contracts more efficiently. Legal contracts can contain lengthy clauses, complex terminology, obligations, conditions, deadlines, and potential areas of concern. Manually reviewing large contracts can require significant time and attention, particularly when users need to identify specific clauses or compare different sections.

The proposed system uses Natural Language Processing, Machine Learning, Generative AI, semantic analysis, and document processing to analyze uploaded contracts. The system extracts text from documents and identifies important sections such as payment terms, termination clauses, confidentiality provisions, obligations, liabilities, renewal conditions, penalties, and other contract-related information.

A clause-classification module categorizes extracted clauses according to their general purpose. A risk-analysis component evaluates clauses against predefined patterns, rules, and configurable risk indicators. Potentially important or unusual clauses can be highlighted within the document, allowing users to focus their attention on specific sections instead of reviewing the entire contract manually.

The system can also provide plain-language explanations of complex clauses and summarize important contract terms. A Generative AI component can explain why a clause was highlighted and identify the specific text supporting the explanation. This makes the system useful for preliminary document review and contract understanding.

Overall, the Legal Contract Analyzer provides an intelligent assistance layer for contract review by combining automated clause extraction, classification, risk highlighting, summarization, and explanation. It is intended to support—not replace—qualified legal professionals. Future enhancements can include multilingual contract analysis, contract comparison, obligation tracking, configurable organizational risk policies, legal-document OCR, improved citation and evidence tracing, and integration with document-management systems.

I. INTRODUCTION

Contracts are important documents used in business, employment, procurement, partnerships, services, finance, and many other activities. They define responsibilities,

rights, obligations, conditions, and other terms agreed upon by parties. Because contracts can contain large amounts of detailed information, reviewing them carefully can be challenging.

Traditional contract review often involves manually reading each section and identifying important clauses. Legal professionals may need to examine definitions, obligations, payment conditions, termination provisions, confidentiality requirements, liability provisions, and other contractual terms. The amount of manual effort can increase significantly when organizations handle large numbers of contracts.

Artificial Intelligence and Natural Language Processing provide opportunities to automate parts of document analysis. NLP techniques can identify entities, clauses, keywords, and relationships within legal text. Machine Learning can classify different types of clauses, while Generative AI can provide simplified explanations and summaries of retrieved contract content.

The **Legal Contract Analyzer** is designed to assist users during preliminary contract review. Users can upload a contract, after which the system extracts and processes its text. The system identifies relevant clauses and highlights potentially important or risk-related sections using configurable rules and AI-based analysis.

The primary objective is to make contract information easier to locate and understand while reducing repetitive document-review effort. Since legal interpretation depends on jurisdiction, contract context, and professional judgment, the system should present its findings as **review assistance rather than legal advice or a definitive determination of legal risk**.

II. LITERATURE SURVEY

Traditional contract-review systems primarily depend on manual document examination and keyword-based search. Legal professionals read contract sections, identify important provisions, and record observations manually. Although this approach can provide detailed review, it can become time-consuming when organizations handle large volumes of documents.

Natural Language Processing has been increasingly applied to legal-document analysis. NLP techniques can extract entities, dates, obligations, parties, amounts, and other relevant information from legal text. Text classification techniques can also be used to categorize clauses according to their purpose.

Machine Learning and deep learning approaches have been explored for legal-text classification and information extraction. Models can learn patterns associated with different clause categories from labeled datasets. However, legal language can vary significantly between contracts, industries, and jurisdictions, making high-quality training data and evaluation important.

Semantic search and embedding-based techniques provide another approach to legal-document analysis. Contract sections can be represented as vector embeddings and compared with predefined clause descriptions, risk indicators, or reference examples.

This can help identify semantically relevant sections even when exact keywords are not present.

Generative AI and Large Language Models have introduced capabilities for summarizing and explaining complex legal text. These models can transform technical language into simpler descriptions and answer questions about retrieved contract content. However, generated explanations may contain unsupported interpretations, making source grounding, evidence tracing, validation, and human review important.

Recent legal AI systems are exploring clause extraction, contract comparison, obligation tracking, risk identification, retrieval-augmented generation, and automated review assistance. These approaches can reduce repetitive work but must address challenges involving accuracy, confidentiality, bias, explainability, and jurisdiction-specific legal interpretation. The proposed **Legal Contract Analyzer** combines clause extraction, classification, configurable risk indicators, source-based highlighting, and Generative AI explanations to support preliminary contract review.

III. SYSTEM ANALYSIS

The **Legal Contract Analyzer** is designed to automatically process contracts and identify clauses that may require closer human review. The system first accepts a contract through a secure document-upload interface. A document-processing module extracts text, page numbers, headings, sections, and other available metadata. OCR can optionally be used for scanned contracts to convert image-based content into searchable text. The text-processing module cleans and divides the document into meaningful clauses or sections while preserving their original locations. An NLP module identifies important entities, obligations, dates, monetary values, parties, conditions, and clause types. A clause-classification engine categorizes provisions such as payment, termination, confidentiality, liability, renewal, and dispute-related clauses. A risk-analysis engine evaluates clauses using configurable rules, patterns, semantic similarity, and predefined risk indicators. Potentially important clauses are highlighted directly in the document and assigned explanatory labels or review categories. A Generative AI module can summarize or explain the highlighted text using the original clause as evidence. A citation layer connects each finding to the relevant document page and clause. Users can review findings, provide feedback, and manually verify or correct classifications. Finally, security, access control, audit logging, and document-retention mechanisms protect confidential contract information.

Existing System

Existing contract-review processes commonly depend on lawyers, compliance teams, contract managers, or business users manually reading documents. Reviewers search through lengthy contracts to identify important provisions and potential areas requiring attention. Keyword searches can help locate specific terms but may fail when the same concept is expressed using different legal language. Manual classification of clauses can also require significant effort. Different reviewers may organize findings using different formats or criteria. Traditional document viewers generally provide limited semantic understanding of contractual language. Some software tools provide templates and basic clause libraries but may require users to manually compare contract text against them. Conventional systems may also provide limited natural-language

explanations of complex provisions. When many contracts need to be reviewed, manual processing can create a significant workload. Scanned documents may require separate OCR processing before they can be searched. Risk identification can also depend heavily on reviewer experience and predefined checklists. These limitations create a need for an AI-assisted system that can automatically locate, classify, and highlight potentially important contract clauses.

Disadvantages of Existing System

- Manual contract review can be time-consuming.
- Large contracts require extensive reading.
- Keyword search may miss semantically similar clauses.
- Clause classification may require manual effort.
- Risk identification can depend heavily on reviewer experience.
- Limited automated highlighting of important provisions.
- Difficult to process large contract collections manually.
- Scanned documents may require separate OCR tools.
- Limited plain-language explanations.
- Limited source-linked AI assistance.
- Difficult to maintain consistent review workflows.
- Manual comparison of clauses can require additional effort.

Proposed System

The proposed **Legal Contract Analyzer — Clause/Risk Highlighter** provides an AI-assisted platform for preliminary contract analysis and review. Users upload contracts through a secure interface, and the system extracts text, page numbers, headings, and other document information. OCR can process scanned documents when required. The extracted text is cleaned and divided into meaningful clauses while preserving document-location information. An NLP module identifies important contractual entities, obligations, conditions, dates, monetary values, and clause categories. A clause-classification engine categorizes provisions such as payment, termination, confidentiality, liability, indemnity, renewal, and dispute-related clauses. A configurable risk-analysis engine evaluates clauses using predefined rules, semantic matching, and organization-specific review criteria. Potentially important clauses are highlighted in the document with appropriate review indicators. The Generative AI module provides concise summaries or plain-language explanations based on the highlighted clause. Each finding is linked to the relevant document page or text segment to improve traceability. Users can review the findings and provide feedback or corrections. The system maintains security and access controls to protect confidential contract documents. Overall, the proposed system reduces repetitive review effort while keeping qualified human reviewers responsible for legal interpretation and final decisions.

Advantages of Proposed System

- Automates preliminary contract analysis.
- Identifies and classifies important clauses.
- Highlights potentially relevant or risk-related sections.
- Supports semantic rather than only keyword-based analysis.

- Provides plain-language clause explanations.
- Reduces repetitive manual searching.
- Supports OCR for scanned documents.
- Provides page-level source references.
- Supports configurable risk indicators.
- Helps organize contract-review findings.
- Supports human review and verification.
- Can process multiple types of contracts.

IV. METHODOLOGY

The methodology begins when a user securely uploads a contract document to the system. The document-processing module identifies the file type and extracts available text, pages, headings, and metadata. If the contract is scanned, an OCR module converts the document images into machine-readable text. The preprocessing module removes unnecessary formatting and divides the contract into meaningful sections or clauses. NLP techniques are then applied to identify entities such as parties, dates, amounts, obligations, conditions, and defined terms. A clause-classification model categorizes the extracted sections according to their general contractual purpose. The risk-analysis module compares clauses against configurable rules, patterns, semantic indicators, and organization-specific review criteria. Potentially important clauses are assigned review labels and highlighted within the original document. The Generative AI module generates concise summaries or plain-language explanations using the identified clause as context. A validation module checks that each generated explanation is connected to the underlying contract text. The citation mechanism provides the document name, page number, section, or source text associated with each finding. Users can manually review findings and provide corrections or feedback. Finally, monitoring, security, access control, audit logging, and data-retention mechanisms are used to maintain system reliability and protect confidential contract information.

SYSTEM ARCHITECTURE

The Legal Contract Analyzer architecture consists of multiple layers for secure document ingestion, legal-text processing, clause classification, risk analysis, and AI-assisted explanation. The User Interface Layer allows authorized users to upload contracts, view analysis results, search clauses, and inspect highlighted findings. The Document Ingestion Layer validates uploaded files and securely stores them for processing. The PDF/Text Processing Layer extracts text, pages, headings, tables, and metadata from contracts. The OCR Layer optionally processes scanned contracts and converts images into machine-readable text. The Text Preprocessing Layer cleans the extracted content and divides it into meaningful contractual clauses while preserving page references. The NLP and Information Extraction Layer identifies parties, dates, amounts, obligations, conditions, and important legal entities. The Clause Classification Layer categorizes provisions according to their general contractual purpose. The Risk Analysis Layer evaluates clauses using configurable rules, semantic matching, and predefined risk indicators. The Generative AI Layer summarizes and explains selected clauses using retrieved contract text as context. The Citation and Evidence Layer connects every finding to the relevant document page, section, or source text. The Review Layer allows users to inspect, verify, correct, and annotate findings. Finally, the Security, Monitoring, and Audit Layer manages authentication,

authorization, encryption, audit logs, document retention, and system-performance monitoring.



V. RESULT AND OUTPUT

Legal Contract Analyzer Clause/Risk Highlighter

Home Upload **Analyze** Results Compare Settings

Search clauses, terms... SR Srijanjitha "Review Smarter. Contract Safer."

Uploaded Document Service_Agreement.pdf 12 pages • 1.2 MB Uploaded on Sep 16, 2026, 10:25 AM 24 Total Clauses 5 High Risk 8 Medium Risk 11 Low Risk Download Report Share New Analysis

Overview Clauses Risk Analysis Summary Ask AI

5 High Risk Clauses Require immediate attention

8 Medium Risk Clauses Review recommended

11 Low Risk Clauses Generally standard

24 Total Clauses Analyzed successfully

Key Findings

- Broad termination rights for either party (Clause 5.2)
- High confidentiality obligations (Clause 6.1)
- Limitation of liability may be unfavorable (Clause 7.1)
- Payment delay penalties (Clause 4.2)
- Standard governing law clause (Clause 8.1)

Contract Summary

This is a Service Agreement between the Client and the Service Provider for a term of 12 months. The agreement covers service delivery, payment terms, confidentiality, limitation of liability, and governing law. There are 5 high-risk clauses that may require negotiation or further review, particularly related to termination rights and liability limitations.

Detected Clause Types

Clause Type	Count
Termination	3
Confidentiality	4
Payment Terms	3
Liability	3
Governing Law	2
Intellectual Property	2
Other	7

Selected Clause Details

Clause 5.2 – Termination High Risk

"Either party may terminate this Agreement for any reason by providing 30 days' written notice to the other party."

Risk Analysis

This clause allows either party to terminate the agreement for any reason with only 30 days' notice, which may create uncertainty and business risk.

Recommendation

Consider adding a specific list of acceptable termination grounds or a longer notice period.

Ask AI about this clause View in Document

Analysis completed successfully! 24 clauses detected • 5 high risk • 8 medium risk • 11 low risk Processed in 28 seconds

VI. CONCLUSION

The Legal Contract Analyzer — Clause/Risk Highlighter provides an AI-assisted approach for analyzing lengthy legal contracts and identifying clauses that may require closer human review. By automating parts of document processing, the system can reduce the time required to locate important contractual provisions.

The system combines Natural Language Processing, clause classification, semantic analysis, configurable risk rules, Generative AI, and document processing to identify and categorize important clauses such as payment terms, termination conditions, confidentiality, liability, renewal, and dispute-resolution provisions.

A key feature of the proposed system is its clause highlighting and evidence-based explanation. Users can view highlighted sections within the original document and receive summaries or explanations connected to the corresponding clause and page. This improves traceability and helps users focus their attention on relevant sections.

Overall, the system can support preliminary contract review by organizing large amounts of contractual information into a more accessible format. It can be useful for legal teams, contract managers, businesses, compliance teams, and organizations that handle multiple contracts.

Future enhancements can include multilingual contract analysis, advanced OCR, contract comparison, obligation and deadline tracking, organization-specific review rules, improved semantic risk detection, automated contract summarization, and integration with document-management systems.

The system should remain an assistive tool rather than a replacement for qualified legal professionals. AI-generated findings and risk indicators should be reviewed by an appropriate legal expert before important contractual decisions are made.

REFERENCES

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
- [3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in

Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in *Blockchain for Smart Cities*, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," *Journal of Machine and Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE)*, Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.

[12] V. N. S. Manaswini, K. K. C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in *Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT)*, Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[13] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment*, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[14] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in *Blockchain for Smart Cities*, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[15] M. M. Rao, M. B. Mohiuddin, P. Srinu, A. Satyanarayana, J. Madhavan and R. D. Kumar, "Carbon-Neutral-Agriculture: Renewable-Energy Powered IoT Systems and Resource Optimization," *2026 3rd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, Chennai, India, 2026, pp. 1-6, doi: 10.1109/RMKMATE69073.2026.11518903.

[16] R. Uma, M. Maggidi, A. R. Ekkati, S. Bandlamudi, C. Madugundi and S. K. Taduri, "Zero-Trust Architecture for Hybrid Cloud Security in Critical Infrastructures," *2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON)*, Lonavala, India, 2026, pp. 1-6, doi: 10.1109/I3CTCON68242.2026.11507458.

[17] VNS Manaswini, Chithanoor Arjun, Yamini Chouhan, & V. Sumalatha. (2026). MB-EFF: A Multi-Branch Environmental Feature Fusion Deep Learning Model for Crop Classification and Recommendation. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(22s), 460–468.

<https://doi.org/10.70917/ijcisim-2026-5453>

- [18] Yamini Chouhan, SIDDOJU DIVYA, VNS Manaswini, & J. Priyanka. (2026). Vision-Threatening Diabetic Retinopathy Detection Using EfficientNetB4: A Two-Stage Transfer Learning Approach. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(22s), 469–476.
<https://doi.org/10.70917/ijcisim-2026-5454>
- [19] Real Time Vehicle Counting and Classification system, **Dr. A.Satyanarayana,et.al** International Research Journal of Modernization in Engineering Technology and Science, Vol-06, Issue-05, 2024 [ISSN: 2582-5208] doi: 10.56726/IRJMETS57804
- [20] Facial Depth Maps: Detecting Sleep Apnea with Deep Learning, **Dr. A.Satyanarayana,et.al** , International Journal of Scientific Development and Research, Vol 9, Issue-5, 2024 [ISSN: 2455-2631]
- [21] Moodchef: Mood Based Smart Recipe Recommender with cooking Assistance, **Dr. A.Satyanarayana,et.al** , International Journal of Progressive Research in Engineering Management and Sciences, Vol-05,Issue-7, July 2025 [ISSN: 2583-5162] doi: 10.58257/IJPREMS42696
- [22] News Article Text Summarization, **Dr. A.Satyanarayana,et.al**, Journal of Emerging Technologies and Innovative Research, Vol-11, Issue -5, 2024, [ISSN: 2349-5162]
- [23] Voice Assistant and gesture controlled virtual mouse using Deep learning Techniques, **Dr. A.Satyanarayana,et.al** , International Research Journal of Modernization in Engineering Technology and Science Vol-6, Issue-5, 2024 [ISSN: 2582-5208] 10.56726/IRJMETS57214