AI-Powered Legal Contract Risk Heatmapper

Yashraj Uttam Lohar[1], Shreehari Pravin Chaudhari[2], Arslaan Sikandar Nadaf[3], Tushar Mahesh Dobale[4]
[1,2,3,4]*Student, Department of AIML Engineering, Dr. Bapuji Salunkhe Institute of Engineering & Technology, Kolhapur*

***Abstract*—The AI-Powered Legal Contract Risk Heat mapper is an artificial intelligence-based system designed to analyse legal contracts and identify clauses that may contain potential risks. Legal contracts often contain complex terminology and lengthy paragraphs, and manual review can be slow and may overlook unfavorable terms. The proposed system uses Natural Language Processing (NLP) and machine learning to extract text from an uploaded contract, divide it into individual clauses, and classify each clause by type and risk level. High-risk clauses are highlighted in red, medium-risk in yellow and low-risk in green, and the system also provides risk scores, plain-English explanations and review questions. The aim is to make preliminary contract review faster, more accessible and easier to understand for students, individuals, startups and small businesses. The system provides preliminary risk identification and informational support only; it does not replace professional legal advice.**

***Index Terms*—Artificial intelligence, clause classification, contract review, legal contract analysis, machine learning, natural language processing, risk heatmap.**

I. INTRODUCTION

Legal contracts are used in employment, business, property, services and procurement. They define responsibilities, payments, confidentiality, ownership, termination, dispute resolution and remedies. However, contracts can be lengthy and use specialised legal language that is difficult for people without legal training to interpret.

A contract may contain a clause that creates an important financial or operational obligation, yet a reader may overlook it in a large document. Manual review takes time and may require professional expertise. Automated tools can extract text and organise information, but basic search alone does not explain why a clause deserves attention. A useful screening system should identify relevant clauses, show the original wording and explain potential concerns clearly.

The proposed AI-Powered Legal Contract Risk Heatmapper is a document-analysis application that accepts a contract, extracts its text, identifies individual clauses and estimates their risk levels using NLP and classification methods. Results are presented through colour-coded highlights, clause categories, risk scores and plain-English explanations. The system is a screening aid and decision-support tool, not a legal opinion, and its output should be checked by a qualified legal professional where appropriate.

A. Need for AI in Contract Risk Analysis

AI support is needed to (1) reduce the time needed to locate clauses that require closer review; (2) make complex language easier to understand through plain-English explanations; (3) classify clauses into categories such as liability, termination, payment and confidentiality; (4) highlight potentially concerning wording within the original contract; (5) summarise risk levels so that users can prioritise review; (6) support consistent preliminary screening across documents; and (7) help users prepare informed questions for a lawyer or contract reviewer.

B. Problem Areas in Conventional Review

Important terms may be overlooked in long documents (clause discovery); non-experts may misunderstand obligations (complex language); manual review may be inconsistent or slow (risk prioritisation); and changes between versions can be difficult to track (document comparison). The proposed system addresses these through clause segmentation and classification, plain-English explanations linked to source text, colour-coded risk indicators and, in future, version comparison.

II. LITERATURE REVIEW

A literature review helps identify established methods, available datasets, common limitations and opportunities for improvement. Relevant areas include document information extraction, legal NLP, clause classification, explainable AI and human-in-the-loop review. Research findings are distinguished here from proposed features, and any model performance claim must be supported by evaluation on suitable data.

A. Legal Document Analysis

Legal document analysis involves identifying and interpreting information in contracts, policies and agreements. Common tasks include extracting text, segmenting sections, recognising clause types, finding defined terms and comparing versions [1], [2]. Automated methods help organise documents, but legal meaning may depend on the full agreement, jurisdiction, definitions and surrounding clauses. A clause should therefore be reviewed in context rather than judged only by isolated keywords.

B. AI and Natural Language Processing

Machine learning (ML) is a branch of AI in which models learn patterns from data rather than relying only on manually written rules. In supervised learning, examples are labelled with known outcomes, such as clause type or a reviewer-assigned risk category. NLP techniques process human language, while transformer models represent contextual relationships between words [3], [4]. Model outputs depend on training data, annotation quality and the match between training examples and the contracts being analysed.

C. Clause Extraction and Segmentation

A contract must be divided into meaningful units before clause-level analysis. Segmentation can use headings, numbering, paragraph boundaries and text patterns. PDF extraction may disrupt reading order or omit text from scanned pages, so page and text-position information should be preserved where possible. Incorrect segmentation can cause the model to miss context or combine unrelated obligations.

D. Legal Clause Classification

Clause classification assigns labels such as payment, termination, liability, indemnity, confidentiality, intellectual property or dispute resolution. Models may use word patterns, conventional machine-learning features or transformer-based representations. Performance should be measured on held-out examples and across different contract types. Ambiguous or unfamiliar clauses should be marked for human review rather than assigned an unjustifiably certain label.

E. Risk Assessment and Scoring

Risk assessment estimates whether a clause may deserve additional attention. A prototype can combine model predictions, clearly defined risk indicators and a score mapped to low, medium or high categories. A numerical score is meaningful only when its definition is documented and evaluated; it should not be presented as an objective measure of legal danger unless validated for that purpose. Context, jurisdiction, bargaining position and the complete contract can change the interpretation.

F. Plain-English Explanation

Explanations can restate a clause in simpler language, identify the wording that triggered a warning and describe a possible practical consequence. A reliable interface should separate the original clause, the model's classification and any generated explanation. Explanations should not invent facts or assert that a clause is unlawful without sufficient legal basis.

G. Explainability, Privacy and Human Review

Legal documents can contain personal, financial, employment and business-confidential information. A contract analysis system should minimise data collection, restrict access and define how uploaded documents are stored or deleted. Explainable outputs and a human-review option are important because incorrect classifications can mislead users, and the application should clearly communicate uncertainty.

H. Conventional and AI-Assisted Approaches

Table I compares conventional review with the AI-assisted approach.

Table I. Conventional and AI-assisted contract review

Aspect	Conventional	AI-assisted
Clause identification	Manual reading and keyword search	NLP-based segmentation and clause classification
Risk review	Reviewer judgement and checklists	Risk indicators with explanations and human verification

Explanation	Legal wording may remain hard to understand	Plain-English summary linked to original clause
Document summary	Manual notes and reports	Dashboard of clause categories and risk levels

III. ARCHITECTURE AND WORKING

A. General Architecture

The application contains connected stages for document upload, text extraction, clause segmentation, NLP classification, risk scoring, explanation generation and visual presentation, as shown in Fig. 1. The precise implementation may vary, but each stage should preserve the link between a result and the original contract text so that users can verify the analysis. Table II lists the core components.

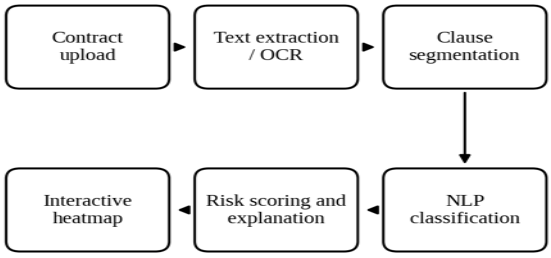

Fig. 1. General architecture of the AI-Powered Legal Contract Risk Heatmapper.

Table II. Core components of the contract risk heatmapper

Component	Function	Examples
Document input	Accept and validate contract files	PDF, DOCX, TXT
Text extraction	Convert content into text	PyMuPDF, python-docx, OCR
Clause processing	Split and prepare text units	Headings, numbering, paragraphs
AI/NLP	Classify clause types and risk indicators	TF-IDF classifier, transformer
Explanation	Describe concern in plain English	Grounded summary, review questions
User interface	Present results for inspection	Interactive highlights, dashboard, export
Security and review	Protect documents, support verification	Access control, deletion, human review

B. Contract Upload and Text Extraction

The user uploads a supported document such as PDF or DOCX. A document-processing component extracts the text [5], [6] and, when needed, an OCR component converts scanned page images into machine-readable text. The system should retain page references and clause boundaries where possible. Password-protected, corrupted or unsupported files should produce a clear error rather than a misleading analysis.

C. Text Preprocessing

Extracted text may contain repeated headers, page numbers, broken line endings, duplicated text or OCR errors. Preprocessing can normalise whitespace, repair common formatting issues and preserve headings, numbering and paragraph boundaries using NLP tooling such as spaCy [7]. The original text remains available for display, because aggressive cleaning can change the meaning of legal wording.

D. Clause Selection and Feature Preparation

Candidate inputs include clause text, heading, surrounding context, document type and structural indicators. Sentence or paragraph boundaries create clause units, and nearby text may be retained to interpret references such as “subject to the foregoing section.” Features are derived only from information in the uploaded document, and unnecessary use of personal information is avoided.

E. Model Selection

A simple baseline is established before adopting a complex model. Possible baselines include keyword/rule-based patterns and TF-IDF with a linear classifier [8]. A transformer model, such as a suitable BERT-family or legal-domain model, may be fine-tuned [9], [10] when a labelled dataset and sufficient computing resources are available. Model choice depends on dataset quality, interpretability, latency and measured performance.

F. Training and Testing

A labelled dataset is divided into training, validation and test sets. Similar clauses or documents from the same source are managed carefully to reduce data leakage. Training fits model parameters, validation supports model selection, and the test set estimates performance on unseen examples. If the dataset is

small or imbalanced, this limitation is reported rather than overstating generalisation.

G. Application Interface

The web interface and API layer that connect the processing pipeline to the user may be implemented with a framework such as FastAPI [11].

IV. EVALUATION METRICS

Table III summarises the measures proposed for each task.

Table III. Model training, testing and evaluation measures

Task	Possible metrics	Interpretation
Clause classification	Precision, recall, F1-score	Class-wise correctness and missed labels
Risk classification	Macro F1, confusion matrix	Performance across risk categories
Explanation review	Human evaluation checklist	Faithfulness, clarity, unsupported claims
Document extraction	Text coverage, error review	Missing or corrupted extracted content

No single metric is sufficient for every classification task. Accuracy can hide poor performance on a less common risk category, so precision, recall and F1-score should also be examined for each class. The project should report its dataset source, label definitions, train/test procedure and errors. Generated explanations should be reviewed separately for factual consistency and faithfulness to the original clause.

V. MAJOR APPLICATIONS

A. Clause Identification and Classification

The system identifies clause units and predicts categories such as payment, termination, confidentiality, liability, indemnity, intellectual property and dispute resolution. Classification organises the contract and lets users focus on specific obligations. The predicted category is displayed with suitable uncertainty and should be correctable when the model makes a mistake.

B. Risk Highlighting and Heatmap

Each clause receives a risk category based on the model and the project's documented risk rules. The interface uses a consistent visual key, as in Fig. 2: green for low, yellow for medium and red for high. The colour indicates a screening result, not a definitive legal judgment. The system shows the reason for the flag and keeps the original wording visible.

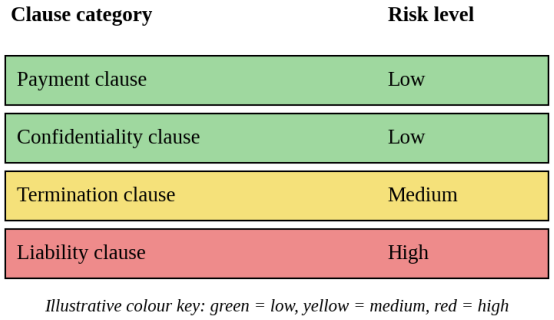

Fig. 2. Clause risk highlighting and review concept (illustrative).

C. Plain-English Explanation and Review Questions

The explanation component restates complex wording in simpler language and describes a possible practical effect. For example, a liability clause without an obvious cap may prompt a question about whether the parties intend to limit financial exposure. Suggestions are framed as questions for discussion, not guaranteed legal remedies or jurisdiction-specific legal advice.

D. Overall Contract Risk Summary

After clause analysis, the dashboard can summarise the number of low-, medium- and high-risk clauses, show category-wise results and list clauses that may deserve priority review. Any overall score must have a transparent calculation and should not imply a validated legal conclusion unless properly tested. The summary links back to each original clause.

E. Report Export and Document Review

The application may provide an exportable summary containing clause references, predicted categories, risk levels, explanations and review questions. It should not change the source contract and should clearly distinguish extracted text from AI-generated commentary. Secure storage, access control and document deletion options are important when contracts contain confidential information.

VI. ADVANTAGES

Faster preliminary review. Automated extraction and clause classification help users locate relevant sections more quickly than searching a long document

manually; this depends on accurate extraction and segmentation and does not remove the need to read the full agreement.

Improved understanding. Plain-English summaries make specialised wording more approachable and remain traceable to the source clause.

Consistent visual prioritisation. A common colour scale and category labels make results easy to scan, provided the colours are explained and not mistaken for legal outcomes.

Support for informed questions. The tool identifies wording that may warrant clarification and suggests questions for a lawyer, employer, supplier or counterparty; it does not tell users to sign or reject a contract.

Reusable and extensible workflow. The modular design can be extended to additional contract types, languages and clause categories as labelled data and evaluation resources become available.

VII. LIMITATIONS AND CHALLENGES

Data availability and label quality. High-quality legal clause datasets may be limited, licensed, imbalanced or unrepresentative of uploaded contracts, and risk labels can reflect reviewer judgement and disagreement.

Model accuracy and context. A clause's meaning can depend on definitions, exceptions, related sections, jurisdiction and the full agreement. False positives and false negatives are possible, so outputs require review.

Privacy and confidentiality. Contracts may contain names, salaries, addresses, signatures and confidential business terms. The system should use secure transmission and storage, limit access, define retention periods and avoid training on uploaded contracts without permission.

Security and file handling. Uploaded files can be malformed or contain unexpected content, so the application should validate file types and sizes, isolate processing and avoid exposing sensitive text in error messages.

Bias and generalisation. A model trained on one jurisdiction, language or contract type may perform poorly on another, and out-of-scope cases should be identified.

Explainability and trust. Users need to know why a clause was flagged; explanations should be grounded in the contract and show uncertainty where appropriate.

Legal accountability and human oversight. Incorrect output could influence important decisions. The application must state its limitations, avoid presenting results as legal advice and encourage qualified review for consequential matters.

Table IV. Advantages and limitations of the proposed system

Potential advantage	Corresponding limitation / risk
Faster clause discovery	Incorrect segmentation may omit or combine text
Accessible explanations	Simplification may lose legal nuance
Consistent risk labels	Model errors and dataset bias remain possible
Prioritised review	Risk scores are not definitive legal conclusions
Automated document processing	Privacy, security and maintenance are required

VIII. PROJECT SCOPE AND OBJECTIVES

A. Problem Statement

Legal contracts often contain long clauses and specialised language that make it difficult for non-experts to identify obligations and potential risks quickly. The problem statement is: how can an AI-assisted application extract clauses from an uploaded contract, classify their type, highlight potentially risky wording and provide plain-English explanations while making uncertainty, data limitations and the need for professional legal review clear? The prototype is a preliminary screening and educational tool; it does not determine enforceability, guarantee the correctness of a risk label, replace a lawyer or make a signing decision.

B. Aim and Scope

The aim is to design and evaluate an AI-assisted application that analyses contract clauses, presents colour-coded risk indicators and generates plain-English explanations linked to the original document. Included in scope are a prototype for selected contract samples, text extraction and clause segmentation, clause categories and risk indicators, plain-English explanations and a heatmap, and evaluation on held-out data with a limitations report. Excluded are

guaranteed legal advice or enforceability decisions, replacement of qualified legal professionals, automatic instructions to sign or reject a contract, guaranteed accuracy across all jurisdictions, and a commercial production service.

C. Objectives

1) Identify suitable legal clause data and verify its source, permitted use, label definitions and limitations.
2) Build a document-processing workflow for supported PDF/DOCX files that preserves clause references.
3) Implement clause segmentation and classify selected clause types and risk categories.
4) Create an interactive heatmap with explanations, evaluation metrics and clear responsible-use safeguards.

IX. CONCLUSION AND FUTURE SCOPE

AI and NLP can support preliminary contract review by extracting text, identifying clause types, organising risk indicators and presenting explanations in a more accessible format. The AI-Powered Legal Contract Risk Heatmapper combines these functions into a proposed workflow that links each risk result to the original clause. This paper described the architecture, preprocessing, clause classification, risk scoring, explanation generation and interface design, together with limitations including data quality, model errors, confidentiality, bias, explainability and the need for human oversight. A risk score is a screening signal rather than a definitive legal assessment. Actual performance must be measured using suitable labelled data, baseline comparisons and held-out evaluation; until implementation and testing are complete, no accuracy or real-world legal effectiveness should be claimed.

Future work includes: fine-tuning a legal-domain transformer on a larger, carefully labelled clause dataset; supporting more contract types, jurisdictions and clause categories after evaluation; adding multilingual processing and improved OCR for scanned documents; comparing contract versions to identify added, removed or modified clauses; linking each explanation statement to the exact source text; adding secure document management, configurable retention and role-based access; providing confidence indicators across document sources; and exploring retrieval-augmented explanations grounded in trusted legal references, with citations and human review.

ACKNOWLEDGMENT

The authors thank the Principal, Dr. S. G. Sapate, and the Head of the Department of Artificial Intelligence and Machine Learning Engineering, Mrs. A. M. Kate, of Dr. Bapuji Salunkhe Institute of Engineering & Technology, Kolhapur, and the internal guide, Ms. P. C. Jasud, for their guidance and encouragement. The authors also acknowledge the researchers and technical organisations whose published work and documentation informed this study.

REFERENCES

[1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. Draft textbook and online materials.

[2] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2008.

[3] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[4] D. Hendrycks, C. Burns, A. Chen, and S. Ball, "CUAD: An expert-annotated NLP dataset for legal contract review," in *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2021. [Online]. Available: https://arxiv.org/abs/2103.06268

[5] J. Govea, I. Ortiz-Garcés, P. Palacios, A. Maldonado Navarro, S. Acurio Del Pino, and W. Villegas-Ch, "Technical evaluation of language models adapted for the automation of legal contracts: Clause extraction, classification, and summarization," *Frontiers in Artificial Intelligence*, vol. 9, Art. no. 1782405, 2026, doi: 10.3389/frai.2026.1782405.

[6] "Legal Indian Contract Clauses Dataset." [Online]. Available: https://share.google/kr3wmLp6NitKidJOO

[7] Singh, A. Joshi, J. Jiang, and H.-Y. Paik, "A survey of classification tasks and approaches

for legal contracts," *Artificial Intelligence Review*, vol. 58, Art. no. 380, 2025, doi: 10.1007/s10462-025-11359-8.

[8] J. Govea, I. Ortiz-Garcés, P. Palacios, A. Maldonado Navarro, S. Acurio Del Pino, and W. Villegas-Ch, "Technical evaluation of language models adapted for the automation of legal contracts: Clause extraction, classification, and summarization," *Frontiers in Artificial Intelligence*, vol. 9, Art. no. 1782405, 2026, doi: 10.3389/frai.2026.1782405.

[9] Chalkidis, M. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos, "LEGAL-BERT: The muppets straight out of law school," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 2898–2904, doi: 10.18653/v1/2020.findings-emnlp.261.

[10] Vaswani *et al*., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.