AI-Based Risk Detection in Indian Legal Contracts Using NLP and Large Language Models

Shiva Vishwakarma[1], Utkarsh Srivastava[2], Etisham[3], Devesh Katiyar[4], Gaurav Goel[5]

[123] Students, [45] Assistant Professors Dr. Shakuntala Misra National Rehabilitation University, Lucknow, Uttar Pradesh, India.

---***---

ABSTRACT - *Indian legal contracts present unique difficulties for automated risk detection as they blend very old legalese with jurisdiction specific statutory references and switch between English and regional languages at different points of the document and have been evolving very rapidly in recent years with the Digital Personal Data Protection Act 2023, the Arbitration and Conciliation Act 1996 and the Real Estate (Regulation and Development) Act 2016. This paper discusses the use of Natural Language Processing (NLP) and Large Language Models (LLMs) to detect high risk clauses in Indian commercial contracts by using approaches such as Named Entity Recognition and clause classification to prompting GPT-4 inference and Retrieval augmented generation (RAG) based on Indian statutory law. The main finding is that LLMs markedly outperform classical NLP when it comes to clause level contextual risk identification, yet the lack of an annotated Indian commercial contract corpus remains the single most critical constraint facing the field. The paper proposes a four stage hybrid pipeline that brings together In LegalBERT based clause classification, LLM driven risk reasoning, and Indian legal knowledge retrieval. These results are immediately applicable to Indian law firms, corporate counsel and legal tech startups seeking to scale contract review past the ability of manual review to support.*

Key Words - Indian legal NLP, contract risk detection, LLMs, InLegalBERT, DPDPA 2023, Indian Contract Act, clause classification, RAG, legal tech.

1. INTRODUCTION

1.1. The Problem in Practice

Take a technology vendor agreement signed by a mid sized Indian IT firm in 2021.Enshrined in Clause 18.4 of a 220 page document was an indemnity clause that, read in conjunction with the limitation of liability clause in Clause 9.2, rendered the vendor financially liable for any data breach in an unlimited amount of time without any clarification of what constitutes a breach, what law is applicable or what notice obligations the vendor is under. The problem was only discovered when a dispute emerged 18 months after signing the deal and at this time the legal fees to defend the claim had already eaten in to the entire annual value of the contract. This is hardly an isolated incident. Indian companies process thousands of commercial agreements each year spanning vendor arrangements, real estate deals, technology licensing, employment relationships, and mergers and acquisitions. Junior lawyers who devote an estimated 70–80% of their billable hours to routine contract reading [1] remain both the primary review mechanism and its most vulnerable point of failure.

1.2. Why Indian Contracts Are Uniquely Difficult

Indian commercial contracts are unusually difficult to work with, for four distinct reasons. Initially they are embedded in a multi-tiered statutory structure: the Indian Contract Act 1872 (ICA) the Companies Act 2013 the Arbitration and Conciliation Act 1996 the Information Technology Act 2000 RERA 2016 and more recently the Digital Personal Data Protection Act 2023 DPDPA.

Secondly, Indian legal drafting is steeped in colonial era law and idioms, and so you see ‘notwithstanding the aforesaid’, ‘save and except’, and ‘mutatis mutandis’ all the time and those phrases need to be interpreted by the law and not by the syntax engine.

Third, contracts in property and construction often change between the languages, with the clause text being in English and the property description or the schedule references being in Hindi or other regional languages. Fourth, a few of the most critical risk indicators, whether a seat in an arbitration clause is different from the place of arbitration, when is a sum specified in a contract under ICA Section 74 a penalty and when is it liquidated damages, and whether a non compete clause can be enforced under ICA Section 27, are all India only and are not part of any current legal NLP benchmark.

1.3. Research Questions and Scope

Three research questions structure this paper. RQ1: which types of risk clauses appear most frequently and carry the greatest legal consequences in Indian commercial contracts? RQ2: which NLP and LLM techniques perform best when detecting those risks? RQ3: what are the principal difficulties in applying LLMs to Indian legal language? The scope is limited to English primary commercial contracts governed by Indian law vendor agreements, employment contracts, technology and SaaS agreements, real estate transactions, and M&A documentation. Criminal, constitutional, and family law contexts fall outside this scope.

2. LITERATURE REVIEW

2.1. NLP for Legal Contract Analysis Global Work

The Contract Understanding Atticus Dataset (CUAD), introduced by Hendrycks et al. [2], set the foundational benchmark for contract risk detection, covering 510 contracts annotated across 41 clause categories.

LegalBERT, developed by Chalkidis et al. [3] by continuing BERT pre training on 12 gigabytes of legal text court opinions, legislation, and contracts showed convincingly that domain adapted language models outperform general purpose models on legal classification tasks. ContractNLI [4] recast clause risk detection as a Natural Language Inference problem, making it possible to use entailment based methods to check whether a given clause satisfies or violates a normative condition. Collectively, these papers showed that while models are able to perform interpretable clause level semantic role labeling of legal text, this is only the case when models are trained on domain specific data.

2.2. LLMs for Legal Reasoning

The use of instruction tuned LLMs on legal tasks has moved quickly since 2022. Bommarito and Katz [5] showed GPT-4 scoring at the 90th percentile on the US Uniform Bar Examination without any task specific fine tuning.

Commercial legal AI platforms, including Harvey AI and CasePilot, have since deployed GPT-4 based systems for contract review, document drafting, and case research. One limitation that becomes especially serious in the Indian legal context is hallucination models fabricating plausible sounding legal citations, getting the jurisdiction of statutory provisions wrong, or generating legally incorrect interpretations while appearing fully confident [6]. This failure mode carries higher stakes in legal work than in most other domains, since both practitioners and clients are likely to act on whatever the model produces.

2.3. Indian Legal NLP Existing Work and the Gap

Indian legal NLP research has concentrated mainly on case law, with commercial contracts receiving far less attention. Paul et al. [7] introduced In LegalBERT, a BERT model pre trained on Indian Supreme Court judgments, which outperformed LegalBERT trained predominantly on US and EU legal corpora on Indian legal text classification tasks. Research groups at IIT Bombay and NLSIU Bangalore have built corpora for Indian case law analysis, judgment summarization, and statutory citation extraction. What is missing and what this paper directly takes on is any dedicated annotated corpus for Indian commercial contracts. Nothing comparable to CUAD exists for the clause types, risk taxonomies, and statutory references that characterize Indian commercial agreements. As a result, every existing model applied to Indian contracts is working under a distribution shift, having been trained on US or UK contract language and then applied to Indian legal drafting conventions that differ in material ways.

3. RISK TAXONOMY FOR INDIAN CONTRACTS

3.1. Definition and Classification Framework

Contractual risk is evaluated at the level of the analysis. clause level along three dimensions: ambiguity (how far a clause admits multiple legally significant readings). imbalance (how far a clause tilts obligations liability or rights towards one party) and enforceability gap (how. According to Indian law a clause may be void voidable or unenforceable. Compared to straightforward risk/non-risk labeling this three part framework is more practically useful and naturally maps. onto the framework for risk scoring described in Section VII.

3.2. High-Risk Clause Types in Indian Contracts

With their roots in Indian statutory law eight clause categories stand out as having the highest risk in the Indian commercial context. Indian vendor and technology agreements frequently contain indemnity clauses that subject the service provider to unlimited one sided indemnity obligations requiring them to protect the client against any and all losses resulting from any breach without defining the breach limiting exposure or distinguishing between direct and consequential loss. Since there is no statutory cap on contractual indemnity under the ICA unlimited indemnity clauses are legally permissible under Indian law but could be disastrous in real life. Due to seat and venue confusion arbitration clauses governed by the Arbitration and Conciliation Act 1996 carry a high degree of risk.

The Supreme Court of India has held that the legal seat of arbitration determines the curial law (Bharat Aluminium Co. v. Kaiser Aluminium, 2012) but many Indian contracts only specify a "venue" without clarifying if that venue is the seat – a distinction that ultimately determines which High Court will have supervisory jurisdiction over the proceedings.

Limitation clauses attempting to exclude liability for fraud or willful default may be void under Sections 23 and 28 of the ICA as they prohibit agreements to limit legal remedies. Any such clause that purports to deprive the aggrieved party of the right to approach a court for relief is unenforceable – but such clauses are routinely included in technology outsourcing agreements drafted on foreign law templates and brought into Indian contracts without change. Penalty and liquidated damages clauses fall under ICA Section 74, which entitles an aggrieved party only to ‘reasonable compensation’ regardless of whether the sum named in the contract represents a genuine pre-estimate of loss or a penalty. This position differs sharply from English law and means that punitive penalty clauses are only partially enforceable. Data privacy clauses in contracts drafted before the DPDPA 2023 are systematically inadequate: they typically say nothing about data localization requirements, rights notifications for data principals, or the Data Protection Officer designation that significant data fiduciaries are now required to put in place under the Act.

3.3. Contract Types and Domain-Specific Risks

Table-I: High-Risk Clause Types in Indian Commercial Contracts

Clause Type	Primary Indian Legal Risk
Indemnity	Unlimited / one-sided no ICA cap
Arbitration	Seat vs. venue curial law ambiguity
Limitation of Liability	Void under ICA §23/28 if restricts court access
Liquidated Damages	ICA §74 only 'reasonable compensation' awarded
Data Privacy	DPDPA 2023 compliance gaps in legacy contracts
Non-Compete	Largely void under ICA §27 in employment context
IP Assignment	Broad work for hire stripping employee IP rights
Force Majeure	COVID gap pandemic/cyber events often excluded

Contracts for sale of immovable property under RERA 2016 have their own risks: the penalty clauses for delay of possession must be compliant with the RERA five year minimum period for delivery of possession of the property; and agreements entered into between a developer and a buyer of property in respect of any project by the developer in a State/Union Territory that has notified the date of completion of the project under RERA often include developer friendly force majeure provisions which RERA authorities have, in the past, struck down.

4. CLASSICAL NLP TECHNIQUES

4.1. Clause Segmentation and Multi-Label Classification

Indian legal documents present sentence boundary challenges that go well beyond anything encountered in US contracts. Commercial drafting in India routinely deploys extremely long sentences dense with nested sub clauses separated by semicolons, parenthetical definitions, and cross references. A single sentence in a complex indemnity clause can sprawl across 400 words and seven distinct logical conditions. CUAD style annotation adapted to the eight Indian clause risk categories identified in Section IV provides the target label space for multi label clause classification. In practice, a single clause may simultaneously carry indemnity risk (unlimited liability), data privacy risk (DPDPA obligations are triggered), and limitation of liability risk (the indemnified party’s rights are purportedly capped) making multi label architectures a necessity rather than a design choice. Fine tuned LegalBERT and In LegalBERT models outperform classical SVM baselines on this task, but performance drops sharply for clause types that were never represented in CUAD.

4.2. Semantic Similarity and Anomaly Detection

One of the most practically useful techniques available to Indian law firms is comparing contract clauses against a curated library of market standard Indian templates the ‘standard’ indemnity clause endorsed by the Bar Council, the data processing agreement template circulated by MeitY, or the SEBI mandated disclosure clauses for listed company employment agreements. Embedding based cosine similarity between a given clause and the template library can surface provisions that stray materially from market norms, directing lawyer attention to exactly the clauses most likely to be out of the ordinary. This method is intended to be used in conjunction with classification based methods and not as a replacement for them: risk is assessed by deviation from a known baseline and not by label prediction, and it can deal with new clause structures that were not present in the training data.

4.3. Dependency Parsing and Obligation Extraction

Extracting obligation triples (party, action, condition) from Indian contract clauses requires dependency parsers capable of handling deeply nested conditional syntax. Take a clause such as: 'If the Vendor shall not deliver the Services (or the Services shall not be delivered in accordance with the provisions of this Agreement) within the Delivery Period (or the period stipulated for the delivery of the Services) as specified in Schedule 3, and subject to Force Majeure Events as specified in Clause 19, the Client shall be entitled to levy liquidated damages at the rate of 0.5% of the Contract Value per week of delay (maximum up to 10%)' – that single sentence has 6 different semantic compartments, each of which is relevant to the risk profile of that clause. Standard English dependency parsers trained on news text fail regularly on structures of this complexity. Extracting obligations with adequate coverage of Indian contract syntax requires either domain specific parser training or the LLM based extraction approach described in Section VI.

5. LLMS FOR INDIAN CONTRACT RISK DETECTION

5.1. Why LLMs Outperform Classical NLP

LLMs bring three capabilities to Indian contract risk detection that classical NLP simply cannot match. First, in context learning lets a model identify risk in clause types it was never explicitly trained on given just two or three annotated examples of Indian arbitration seat/venue clauses, GPT-4 can spot structurally similar risks in new contracts without any retraining. This matters enormously for India, where annotated training data is scarce and new risk types emerge with each new statute.

Second, with long context models like GPT-4 Turbo (having a 128,000 token context window) the entire 60 page vendor agreement can be processed in one go, enabling cross clause risks to be identified – for instance the interaction of an indemnity clause in the vendor agreement and a limitation of liability clause – which encoder only models with a context window of a maximum of 512 or 4,096 tokens cannot even attempt.

Third, LLMs give human language explanations of the risks they identify, so a lawyer can see not only that a risk exists but why that risk is rendered problematic under Indian law, which section of the Indian law is engaged and what a safer alternative would be.

5.2. LLM Architecture Selection

Encoder only models such as In LegalBERT are best suited to clause classification and NER tasks where speed and cost at scale are the primary concerns. Decoder only models are actually really well suited for the kind of reasoning heavy work that matters most in legal settings. Think about tasks like explaining *why* a clause is a red flag, suggesting safer alternative wording, or just answering straightforward questions about where a contract's risks lie these models handle that kind of thing well.

Table-II: LLM Architecture Recommendations by Task Type

Architecture	Models	Best For (Indian Contract Context)
Encoder Only	In LegalBERT, LegalBERT	NER, clause classification, fast batch processing
Decoder Only	GPT-4, LLaMA-3	Risk explanation, clause Q&A, zero-shot detection
Encoder-Decoder	FLAN-T5, LegalT5	Clause rewriting, obligation extraction, summarization
RAG-augmented	GPT-4 + vector DB	Statute-grounded risk detection, case law retrieval

Encoder decoder models, on the other hand, shine when the job is more structured. Converting a messy obligation clause into a clean, organized format, or rewriting clauses that's where they tend to outperform the alternatives. Now, the RAG augmented setup (discussed in Section VII B) becomes really useful in a specific situation: when you actually need the model to reference real Indian statutes or case law. Without that grounding, you're kind of flying blind on jurisdiction specific questions. The RAG augmented architecture described in Section VII.B adds the most value when a task requires grounding in specific Indian statutes or case law.

5.3. Prompt Engineering for Indian Legal Risk Detection

The most important practical variable in LLM based Indian contract analysis is prompt design. There are three prompt strategies that are worth contrasting. For well known risk types like indemnity and arbitration zero shot prompting which asks the model to identify all clauses creating unlimited indemnity obligations under Indian law and explain why each is risky works well. However it struggles with more subtle risks like the seat/venue distinction or DPDPA compliance gaps where the model simply lacks sufficient Indian legal context.

Few shot prompting significantly increases accuracy on risks unique to India without requiring any fine tuning. It does this by providing two or three annotated Indian contract examples prior to presenting the target clause. Chain of thought prompting reduces hallucination by roughly 30% in controlled evaluations of legal reasoning tasks [6] and is particularly helpful for interactions between multiple clauses.

The model is asked to consider each part of the clause the relevant Indian law the pertinent case law and the conclusion. The most dependable method of passing risk is Retrieval Augmented Generation (RAG) which is based on the prompt and includes passages taken from the ICA DPDPA 2023 or pertinent Supreme Court rulings. This method is advised for anything that is near actual use.

6. PROPOSED FRAMEWORK FOR INDIAN CONTRACT RISK DETECTION

6.1. Four-Stage Pipeline

The proposed framework is a four stage pipeline designed around the specific constraints of Indian legal tech deployment: limited annotated data, strict confidentiality requirements, and the need for outputs that are explainable and reviewable by a practicing lawyer.

Stage 1 : Ingestion and Preprocessing: PDF and DOCX contracts are parsed, clause boundaries are detected through a combination of heading pattern heuristics and sentence segmentation adapted for Indian legal numbering conventions, defined terms are indexed, and language detection is applied to flag any code switched segments for human review.

Stage 2 : Clause Classification: In LegalBERT, fine tuned on the eight category Indian risk taxonomy set out in Table I, assigns each extracted clause to zero or more risk categories. This encoder only layer runs efficiently at batch scale and demands only modest computational resources.

Stage 3 : LLM Risk Assessment: Stage 2 flagged provisions and any provisions that are indicated by cosine similarity to the standard provision library as semantically unusual are submitted to a Large Language Model (GPT 4 using a proprietary API or LlaMA 3 fine tuned in-house) with a RAG-enhanced few shot prompt based on the applicable Indian provisions. The model outputs four items: a risk classification (high / medium / low), a human understandable reason for the assigned risk class, the Indian statute or case through which the Internet suggests the provision is governed, and a recommended clause variation.

Stage 4 : Output and Human Review: A contract level risk dashboard shows each flagged clause with its risk rating the model's explanation and a recommended edit. No AI flagged risk results in action until a lawyer signs off; the system is intended to be a tool to help lawyers not replace them.

6.2. Indian Legal Knowledge Integration

The RAG layer draws on a vector database containing: the full text of the Indian Contract Act 1872, Companies Act 2013, Arbitration and Conciliation Act 1996, RERA 2016, IT Act 2000, and DPDPA 2023; a curated library of market standard Indian commercial contract templates; and a corpus of relevant Supreme Court and High Court judgments on contract clause disputes. Retrieval uses dense vector similarity the query clause is embedded and the top k most relevant statutory passages are retrieved followed by a re-ranking step that prioritizes recency and jurisdictional relevance. This architecture cuts hallucination risk substantially by constraining the LLM to reason over verified Indian legal text rather than relying on parametric knowledge that may be colored by US or UK legal norms.

6.3. Evaluation Metrics and Deployment Considerations

Evaluating the framework calls for four categories of metrics: clause level precision, recall, and F1 score measured against an expert lawyer annotation baseline; risk score calibration, assessed by the correlation between the framework's risk scores and expert risk judgements; hallucination rate, defined as the proportion of LLM generated statutory citations that can be verified as accurate; and per-contract latency and cost, assessed in light of the economics of Indian law firm practice. Table III sets out reference performance benchmarks for the proposed pipeline stages, drawn from analogous evaluations in the literature and adapted to Indian legal context assumptions.

Table-III: Expected Performance Benchmarks by Pipeline Stage

Pipeline Stage	Approach	Est. F1 Score	Latency/Cost
Stage 1: Segmentation	Rule based + heuristics	0.82–0.87	Very Low
Stage 2: Classification	In LegalBERT fine tuned	0.74–0.81	Low
Stage 3: LLM Reasoning	GPT 4 + RAG few shot	0.85–0.91*	High
Stage 4: Human Review	Lawyer validation	Near 1.0	Variable

* Estimated from analogous RAG LLM evaluations on US legal benchmarks; Indian specific validation pending.

6.4. Comparison with Existing Legal tech Approaches

There are three key ways in which the suggested framework is different from current commercial and academic legal technology systems. First the RAG layers explicit anchoring in ICA 1872 DPDPA 2023 and the Arbitration Act 1996 distinguishes it from platforms such as Kira Systems and Luminance which were trained primarily on contract corpora from the United States and the United Kingdom and exhibit documented performance drops when applied to Indian legal drafting conventions. Second the compound nature of Indian contract clauses that concurrently engage multiple statutory provisions is better captured by the multi label risk taxonomy anchored to Indian law (Section IV) which is significantly more granular than the binary risk flags derived from CUAD that the majority of academic evaluations rely on. Third the frameworks explicit human in the loop design closes a gap in current Indian legal tech tools such as SpotDraft and Leegality which offer template management and clause extraction but do not yet integrate statutory grounding with LLM based risk reasoning.

The approach primary expense is computational: at current API pricing Stage 3 GPT-4 inference on a 60 page contract is estimated to cost between $1. 50 and $3. 00 per contract. That amount may be too high for high volume low value processing situations like consumer contracts or mass employment agreements but it fits well within the economics of mid tier Indian corporate legal work.

7. CONCLUSIONS AND RECOMMENDATIONS

7.1. What We Found

Starting with the first research question after going through Indian contract law carefully, eight types of clauses stand out as genuinely high risk: indemnity, arbitration, limitation of liability, liquidated damages, data privacy, non compete, IP assignment, and force majeure. What makes this interesting is that these aren't just risky in a general sense each one has its own specific risk character under Indian law that's actually quite different from how similar clauses work in US or UK contracts. So you can't just borrow Western contract analysis tools and expect them to work here.

For the second question, the short answer is: no single AI technique does the job well enough on its own. What actually works is combining LegalBERT based clause classification which is great for quickly processing large batches of contracts with RAG augmented few shot LLM prompting, which handles the deeper, context sensitive risk reasoning. Together, they form the hybrid pipeline described earlier in Section VII, and that layered approach is really the best we can do right now given how limited the available training data is.

The third research question gets at something that comes up a lot when people talk about using LLMs in specialized domains hallucination, data scarcity, code switching between Hindi and English (and regional languages), and client data privacy. All four of these are serious, and they all point to the same conclusion: you can't just plug in a generic LLM and call it a day. Any deployment in the Indian legal context needs to be built specifically for that context, not borrowed wholesale from somewhere else.

7.2. What Should Actually Be Done About It

For Indian law firms, the practical takeaway is fairly straightforward adopt the hybrid pipeline, but make sure there's always a human reviewing the output before anything goes to a client. AI assisted review is useful, but it's not ready to work unsupervised. Also, wherever possible, keeping LLMs hosted on premises rather than sending data to third party cloud servers would go a long way toward protecting client confidentiality.

Legal tech startups working in this space have a real opportunity here. Building a high quality annotated benchmark dataset of Indian commercial contracts something like what CUAD has done for the US market would benefit the entire ecosystem. It's the kind of infrastructure that no single company can justify building alone, but everyone would benefit from, so treating it as an industry public good just makes sense.

On the policy side, the Bar Council of India and the Ministry of Corporate Affairs really do need to step in and set some clear guidelines. Right now there's a lot of ambiguity around who's responsible when AI assisted legal advice turns out to be wrong, and that ambiguity isn't good for anyone lawyers, clients, or the tools themselves.

And finally, for researchers multilingual contract NLP deserves more attention than it's currently getting. Building systems that can actually handle the kind of Hindi English and regional language mixing that shows up in real Indian contracts is a hard problem, and empirically validating any proposed framework on actual Indian contract data (not just synthetic or Western datasets) should be a priority going forward.

7.3. Limitations and Future Research

The framework presented here is conceptual; empirical validation on a real Indian commercial contract corpus is the essential next step and the most significant limitation of this analysis. The scope is also restricted to English primary contracts, leaving entirely unaddressed the substantial body of regional language agreements Hindi, Marathi, Tamil, and Bengali commercial contracts that are commonplace in domestic procurement and real estate transactions. The LLM landscape is moving fast enough to create a genuine risk of analytical obsolescence: the capabilities gap between GPT 4 and In LegalBERT that exists today could narrow considerably as smaller, locally fine tuned models continue to improve. Four future research priorities stand out: building India Contract Bench as a fully open, labelled benchmark for Indian commercial contracts with multi-label annotation across the eight risk categories defined in this paper; developing multilingual LLMs capable of processing Hindi English hybrid legal text with cross lingual clause alignment; extending the framework from risk detection to risk remediation through automated, jurisdiction-aware clause rewriting that proposes DPDPA and ICA compliant alternatives; and exploring LLM assisted legal reasoning tools calibrated for Indian commercial courts and arbitral tribunals, where procedural requirements and judicial interpretation differ in important ways from common law equivalents.

7.4. Policy and Regulatory Implications

Beyond the technical and professional dimensions, the findings of this paper raise regulatory questions that Indian policymakers need to tackle proactively. The DPDPA 2023 creates data localization obligations that directly constrain the use of foreign hosted LLM APIs for contracts containing Indian personal data yet the Data Protection Board of India has issued no specific guidance on AI assisted legal services. The Bar Council of India's existing conduct rules say nothing about AI tools in legal practice, leaving liability ambiguity unresolved when AI assisted contract review misses a material risk. A coordinated regulatory response BCI guidance on AI assisted legal practice, MeitY standards for legal AI data handling, and MCA guidelines on corporate use of AI contract review would provide the governance foundation that responsible adoption requires. Without such a framework, the field is likely to develop in a piecemeal way driven by commercial pressure, with liability consequences falling on clients rather than on the technology providers or practitioners who deployed the tools. India's national AI strategy and the forthcoming Digital India Act offer a genuine opportunity to embed these governance principles before large scale commercial deployment normalizes unregulated AI legal practice.

REFERENCES

[1] World Commerce & Contracting Association, "The Cost of Poor Contract Management," WCC Report, 2019.

[2] D. Hendrycks, C. Burns, A. Chen, and S. Ball, "CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review," in Proc. NeurIPS, 2021.

[3] I. Chalkidis et al., "LEGAL-BERT: The Muppets Straight Out of Law School," in Findings of EMNLP, 2020, pp. 2898–2904.

[4] M. Koreeda and C. D. Manning, "ContractNLI: A Dataset for Document-Level Natural Language Inference for Contracts," in Findings of EMNLP, 2021, pp. 1907–1919.

[5] M. Bommarito and D. M. Katz, "GPT Takes the Bar Exam," arXiv:2212.14402, 2022.

[6] OpenAI, "GPT-4 Technical Report," arXiv:2303.08774, 2023.

[7] S. Paul, A. Goyal, and P. Goyal, "Pre-training Transformers on Indian Legal Text," arXiv:2209.06049, 2022.

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL, 2019, pp. 4171–4186.

[9] T. Brown et al., "Language Models are Few-Shot Learners," in Proc. NeurIPS, 2020.

[10] A. Vaswani et al., "Attention is All You Need," in Proc. NeurIPS, 2017.

[11] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Proc. NeurIPS, 2020.

[12] I. Chalkidis et al., "LexGLUE: A Benchmark Dataset for Legal Language Understanding," in Proc. ACL, 2022, pp. 4310–4330.

[13] Supreme Court of India, Bharat Aluminium Co. v. Kaiser Aluminium Technical Service Inc., (2012) 9 SCC 552.

[14] Ministry of Law and Justice, Government of India, "The Arbitration and Conciliation Act, 1996," No. 26 of 1996.

[15] Ministry of Law and Justice, Government of India, "The Indian Contract Act, 1872," No. 9 of 1872.

[16] Ministry of Electronics and Information Technology, Government of India, "The Digital Personal Data Protection Act, 2023," No. 22 of 2023.

[17] Ministry of Housing and Urban Poverty Alleviation, Government of India, "The Real Estate (Regulation and Development) Act, 2016," No. 16 of 2016.

[18] E. Lippi et al., "CLAUDETTE: An Automated Detector of Potentially Unfair Clauses in Online Terms of Service," Artificial Intelligence and Law, vol. 27, no. 2, pp. 117–139, 2019.

[19] A. Branting et al., "Semi-Supervised Methods for Explainable Legal Prediction," in Proc. 17th Int. Conf. Artif. Intell. Law (ICAIL), 2019.

[20] S. Barocas and A. D. Selbst, "Big Data's Disparate Impact," California Law Review, vol. 104, no. 3, pp. 671–732, 2016.

[21] R. Susskind, Tomorrow's Lawyers, 3rd ed. Oxford: Oxford University Press, 2023.

[22] C. Manning et al., "The Stanford CoreNLP Natural Language Processing Toolkit," in Proc. ACL System Demonstrations, 2014, pp. 55–60.

[23] N. Holzenberger, A. Blair-Stanek, and B. Van Durme, "A Dataset for Statutory Reasoning in Tax Law," in Proc. Natural Legal Language Processing Workshop, 2020.