

# **Market Survey of LLM Applications and Resources for Commercial Contract Analysis and Generation**

## **1. Executive Summary**

The landscape of legal technology is undergoing a significant transformation with the advent of Large Language Models (LLMs). This report surveys the current market for LLM-based applications designed to assist legal professionals in commercial contract generation and risk evaluation. The analysis reveals a diverse ecosystem of commercial platforms offering features ranging from automated contract review and risk assessment to AI-assisted drafting. Furthermore, the report identifies publicly available datasets that can be utilized for training LLMs in the legal domain, with a particular focus on commercial contracts. Several pre-trained language models, both general-purpose and domain-specific, are examined for their potential application in this context, along with open-source tools and libraries that can facilitate the development of such applications. The findings suggest a mature and competitive market with significant opportunities for innovation, provided developers carefully consider data quality, model selection, and user integration.

## **2. Introduction: The Need for LLM-Powered Contract Solutions**

Commercial contracts form the bedrock of business operations, outlining obligations, rights, and responsibilities between parties. Traditionally, the generation and evaluation of these contracts have been labor-intensive processes for lawyers, demanding meticulous review to ensure accuracy, identify potential risks, and maintain compliance with legal standards. This often involves significant time investment and carries the inherent risk of human error, especially when dealing with large volumes of complex documents. Large Language Models, with their advanced natural language understanding and generation capabilities, present a compelling solution to these challenges. By leveraging these AI-powered tools, legal professionals can potentially streamline their workflows, enhance accuracy, and focus on higher-value strategic tasks. The increasing recognition of AI's value within the legal sector is evident, with a considerable proportion of lawyers already employing generative AI tools for various purposes<sup>1</sup>. This growing acceptance indicates a favorable environment for the adoption of specialized LLM-based contract solutions. This report aims to provide a comprehensive survey of the existing market, available resources, and best practices to inform the development of such an application.

## **3. Market Survey of Existing LLM Applications for Legal Contracts:**

- **3.1 Commercial Platforms for Contract Generation and Risk Evaluation:**

A multitude of commercial platforms are currently available, offering LLM-powered features to assist legal professionals in managing commercial contracts throughout their lifecycle. These platforms cater to various needs, from focused AI-driven contract review to comprehensive contract lifecycle management solutions.

**Robin AI** stands out as a platform centered on accelerating contract review and analysis. It offers features designed to summarize legal documents, meticulously check defined terms within contracts, and even redraft clauses as needed <sup>2</sup>, B1<sup>2</sup>. The platform claims to achieve an impressive 80% faster contract review process, a significant efficiency gain that appeals directly to the time-sensitive nature of legal work <sup>2</sup>. This potential for substantial time savings underscores a key advantage of adopting LLM-based tools in the legal domain.

**Luminance** presents itself as an end-to-end AI platform designed to enhance every interaction a business has with its contracts. This encompasses the entire process from initial contract generation through negotiation and even post-execution analysis <sup>3</sup>, B2<sup>4</sup>. A notable feature is its AI-powered negotiation capability, which functions like a legal "spellchecker," providing a visual "traffic light analysis" to immediately highlight terms that deviate from previously agreed-upon language <sup>3</sup>. This intuitive visualization of risk allows legal teams to quickly identify potential areas of concern. Furthermore, Luminance aims to empower non-legal teams by enabling them to generate contracts using approved templates, thereby ensuring adherence to internal standards <sup>3</sup>.

**LegalFly** offers a platform focused on rapid contract review, capable of analyzing hundreds of contracts in mere seconds <sup>5</sup>. It goes beyond simple review by providing automatic redrafting capabilities to ensure clauses are compliant with relevant regulations <sup>5</sup>. A particularly powerful feature of LegalFly is its custom agent functionality, which allows legal teams to train AI models using their own specific legal and financial documents <sup>5</sup>. This ability to tailor the AI to a firm's unique requirements and workflows offers a significant advantage in terms of accuracy and relevance.

**Spellbook** distinguishes itself through its seamless integration with commonly used word processors, leveraging the power of GPT-4o and other large language models <sup>5</sup>. This integration allows for AI-driven suggestions directly within the familiar environment of tools like Microsoft Word, assisting lawyers in both drafting new legal documents and reviewing existing ones <sup>5</sup>. This ease of integration within established workflows is crucial for facilitating adoption among legal professionals.

Other platforms like **Evisort** focus on automating the entire contract lifecycle, from the initial drafting stage all the way through to renewal, incorporating

intelligent search functionalities and comprehensive reporting features <sup>5</sup>. **Kira Systems**, now part of Litera, specializes in the extraction and in-depth analysis of contract provisions using AI, proving particularly useful for due diligence processes and large-scale document review projects <sup>5</sup>. **ContractPodAi** offers an end-to-end solution for managing contracts, integrating AI features for tasks such as clause identification, risk assessment, and ensuring compliance with relevant regulations <sup>5</sup>, B3<sup>6</sup>. Notably, ContractPodAi emphasizes ethical considerations and incorporates guardrails into its AI-driven processes, reflecting a growing awareness of the importance of responsible AI development in the legal field <sup>6</sup>. **BlackBoiler** automates the analysis, review, and revision of contracts by comparing incoming agreements against an organization's established playbook, offering automated redlines and suggested edits <sup>5</sup>. Its integration with popular document management systems like iManage and NetDocuments further streamlines legal workflows <sup>5</sup>. **DocJuris** provides a platform specifically designed for contract review and negotiation, aiming to accelerate the negotiation process while maintaining compliance <sup>5</sup>, B3<sup>4</sup>. It achieves this by identifying deviations from standard terms and offering negotiation playbooks to guide decision-making <sup>5</sup>, B3<sup>4</sup>.

**CobbleStone Software** offers VISDOM AI, a powerful artificial intelligence backed by machine learning, with a suite of features relevant to contract management <sup>7</sup>. This includes IntelliDraft.ai, a generative AI tool for drafting new clauses, and auto-redlining capabilities to automatically insert pre-approved clauses <sup>7</sup>. VISDOM AI also provides risk assessment mapping to pinpoint potential risks within contracts by analyzing the document text <sup>7</sup>. **Icertis** provides a Risk Assessment Copilot, built upon its ExploreAI Service, which leverages AI to compare contract language against a company's predefined risk parameters, potentially reducing review cycle times by up to 40% <sup>8</sup>. **LexCheck** focuses on AI-powered contract risk assessment and standardization, claiming a rapid two-to-three-minute turnaround for a full contract review <sup>9</sup>. **Ivo** offers AI contract review software equipped with features for flagging high-risk provisions, generating structured issue lists, and integrating with user-defined playbooks <sup>10</sup>. The platform allows for the creation of completely customizable playbooks, enabling firms to tailor the risk assessment process to their specific standards <sup>10</sup>. Broader legal technology platforms like **HighQ by Thomson Reuters** also incorporate contract lifecycle management features, offering secure document sharing, project management, and workflow automation capabilities relevant to contract handling <sup>11</sup>. **Percipient** takes a unique approach by combining AI-powered contract review with the critical element of human expert oversight <sup>4</sup>. This hybrid model acknowledges the limitations of AI in fully grasping complex

legal contexts and ensures a higher level of accuracy and nuanced judgment.

**Juro** offers an AI-powered contract automation platform encompassing contract creation, review, and management within a single, integrated environment <sup>4</sup>.

Finally, **Conga** provides a contract management platform designed to automate the entire contract lifecycle, from the initial creation to ongoing compliance, focusing on streamlining workflows and automating approval processes <sup>4</sup>.

- **3.2 Feature Comparison and Functionality Analysis:**

The features offered by these commercial platforms can be broadly categorized into functionalities for contract generation and risk evaluation.

**Contract Generation Features:** Many platforms offer AI-assisted drafting capabilities, often leveraging templates and pre-approved clause libraries to expedite the contract creation process <sup>7</sup>, B4<sup>3</sup>, B2<sup>6</sup>. Some, like CobbleStone's IntelliDraft.ai and Icertis's Copilot, provide the ability to generate new clauses based on precedents, user input, or through natural language queries <sup>7</sup>, B4<sup>8</sup>.

Automated redlining features, as seen in CobbleStone and Icertis, can automatically incorporate clauses from approved libraries, ensuring consistency and adherence to organizational standards <sup>7</sup>, B4<sup>8</sup>.

**Risk Evaluation Features:** A significant number of platforms focus on risk evaluation. This often includes automated clause identification and categorization based on predefined risk levels <sup>7</sup>. Features like risk scoring and visualization, offered by CobbleStone, Icertis, LexCheck, and ContractPodAi, provide an intuitive way to understand the potential risks associated with a contract at both the overall and individual clause levels <sup>7</sup>, B4<sup>8</sup>, B5<sup>9</sup>, B6<sup>13</sup>, B7<sup>6</sup>. Many platforms, including BlackBoiler and LexCheck, compare contracts against internal playbooks and market standards to identify deviations and potential issues <sup>5</sup>. The ability to identify missing or non-standard clauses is another crucial risk evaluation feature offered by platforms like Luminance and CobbleStone <sup>12</sup>, B2<sup>5</sup>, B4<sup>8</sup>, B5<sup>3</sup>. Platforms such as LegalFly and DocJuris focus on detecting inconsistencies and errors within contract drafts <sup>5</sup>. Several tools, including Robin AI and Kira Systems, offer specific features to support due diligence processes <sup>1</sup>, B3<sup>2</sup>. Natural language querying for risk insights, as provided by Icertis, allows users to ask specific questions about potential risks within their contracts <sup>8</sup>. Some platforms, like CobbleStone, even incorporate sentiment analysis to understand the tone and potential implications of the contract language <sup>7</sup>.

- **3.3 Strengths and Limitations of Current Offerings:**

Current LLM applications for legal contracts offer several compelling strengths. They significantly increase efficiency and speed in contract review and analysis, allowing legal teams to process documents much faster than manual methods <sup>1</sup>. These tools also improve accuracy and consistency in identifying risks and

extracting key information from complex legal texts <sup>1</sup>. By automating routine tasks, these platforms can lead to substantial cost reductions for law firms and legal departments <sup>1</sup>. The enhanced due diligence capabilities offered by many of these tools are particularly valuable during mergers, acquisitions, and other significant transactions <sup>1</sup>. Furthermore, the consistent application of AI-driven analysis can contribute to better compliance and a reduction in overall legal risks <sup>2</sup>. Ultimately, by taking over repetitive and time-consuming tasks, these LLM applications empower legal professionals to focus their valuable time and expertise on more strategic and complex matters <sup>2</sup>.

Despite these significant advantages, current LLM offerings also have limitations. They often lack the nuanced "situational awareness" and comprehensive legal context that experienced human attorneys possess <sup>12</sup>. Consequently, human oversight and thorough validation remain crucial, especially when dealing with high-stakes legal documents <sup>1</sup>. The potential for errors or missed critical nuances exists if the AI models are not adequately trained on relevant data or properly customized to specific legal requirements <sup>16</sup>. Highly complex or non-standard contracts can also pose challenges for current LLM applications <sup>14</sup>. The effectiveness of these tools is heavily dependent on the quality and relevance of the data used to train them <sup>14</sup>. Finally, there is the potential for LLMs to generate inaccurate or nonsensical information, a phenomenon sometimes referred to as "hallucination" <sup>17</sup>.

#### 4. Publicly Available Datasets for Training Legal LLMs on Commercial Contracts:

- **4.1 General-Purpose Datasets with Legal Text:**

Several large, publicly available datasets, while not exclusively focused on legal text, contain a significant amount of textual data that may include legal documents and information relevant to commercial contracts.

**Common Crawl** is a massive repository comprising terabytes of raw web data extracted from billions of web pages <sup>20</sup>. This dataset has been used to train numerous prominent large language models, including GPT-3 and LLaMA <sup>20</sup>. While it undoubtedly contains a vast amount of legal information, including contracts that are publicly available on the internet, extracting and filtering this specific data for training a specialized model would require substantial effort <sup>20</sup>. The sheer volume of Common Crawl is a significant advantage, but its general nature necessitates careful curation to ensure the data is relevant and of high quality for legal applications.

**Wikipedia**, a collaborative online encyclopedia, contains a wide array of articles covering diverse topics, including law and legal concepts <sup>20</sup>. The cleaned text data from Wikipedia, available in various languages, amounts to a substantial corpus

that has been used to train models like RoBERTA and LLaMA<sup>20</sup>. Similar to Common Crawl, while Wikipedia provides a broad overview of legal topics, its legal content might lack the depth and structured format required for training highly specialized models for commercial contract analysis. While it can provide foundational knowledge of legal concepts, its breadth might compromise the depth needed in specific areas like commercial contracts.

**The Pile** is an 800 GB corpus curated from 22 diverse datasets, primarily from academic and professional sources<sup>20</sup>. This dataset was instrumental in training models like GPT-Neo and LLaMA<sup>20</sup>. The diversity of The Pile could potentially include legal datasets relevant to commercial contracts. However, the specific content would need to be investigated to determine its suitability and quality for this purpose. The curated nature of The Pile makes it potentially more valuable than raw web crawls, but its specific legal content needs to be assessed.

The **C4 (Colossal Clean Crawled Corpus)** is a 750 GB English corpus derived from Common Crawl, with a focus on extracting natural language data and removing gibberish text, along with heavy deduplication to enhance its quality<sup>20</sup>. Language models like MPT-7B and T5 have been pre-trained using C4<sup>20</sup>. While C4 offers the advantage of cleaner data compared to raw Common Crawl, its overall suitability for training legal contract models depends on the presence and quality of relevant data within the corpus. The improved quality of C4 over raw Common Crawl makes it a better starting point, but the focus remains general web text.

Other general-purpose datasets like **BookCorpus**, a collection of unpublished books<sup>20</sup>, are less likely to contain significant amounts of legal contract data.

**ROOTS**, a 1.6TB multilingual dataset<sup>20</sup>, could be valuable if the application requires multilingual contract analysis. **Red Pajama**, an open-source effort to replicate the LLaMA training dataset<sup>20</sup>, shares similar characteristics and considerations as the original LLaMA dataset in terms of legal content.

- **4.2 Specialized Legal Contract Datasets (e.g., LegalBench, CUAD):**

In addition to general-purpose datasets, there are specialized datasets specifically designed for evaluating legal reasoning in LLMs, some of which include tasks related to contracts.

**LegalBench** is an ongoing open science initiative focused on collaboratively curating tasks for evaluating legal reasoning in English LLMs<sup>22</sup>, B10<sup>22</sup>. The benchmark currently comprises 162 tasks gathered from 40 contributors, spanning various legal domains and text types, including contracts<sup>22</sup>. While LegalBench includes contract-related tasks, it is primarily intended for benchmarking and evaluating model performance rather than extensive fine-tuning, as its "train" splits are generally small<sup>22</sup>. Nevertheless, it serves as a



valuable resource for assessing how well LLMs perform on specific legal contract-related reasoning tasks.

The **CUAD (Contract Understanding Atticus Dataset)** is a corpus specifically designed for research and development in legal contract review <sup>26</sup>. It contains over 13,000 labels across 510 commercial legal contracts, with manual annotations identifying 41 categories of important clauses that lawyers typically look for during contract review in corporate transactions <sup>26</sup>. The dataset is curated and maintained by The Atticus Project, Inc. and is available on Hugging Face Datasets <sup>26</sup>. The contracts within CUAD are in English and cover a range of commercial agreements relevant to mergers and acquisitions, investments, and initial public offerings <sup>26</sup>. The dataset is released under the CC BY 4.0 license, allowing for research and development use <sup>29</sup>. Given its focus on commercial contracts and the availability of expert annotations, CUAD appears to be a highly suitable dataset for training and evaluating LLMs for tasks such as legal contract review, clause identification, and risk assessment.

#### **Contract Understanding Atticus Dataset (CUAD) Details:**

Feature	Description
Size	510 commercial legal contracts, 13,000+ labels
Content	Manually labeled clauses across 41 categories relevant to corporate transactions
Language	English
License	CC BY 4.0
Suitability	Highly suitable for training and evaluating models for legal contract review, clause identification, and risk assessment.
Download	Available on Hugging Face Datasets: <a href="https://huggingface.co/datasets/theatticusproject/cuad">https://huggingface.co/datasets/theatticusproject/cuad</a>

The **\*\*Atticus Open Contract Dataset (AOK)\*\***, mentioned on Kaggle, is likely related to CUAD and also represents a collection of legal contracts with rich expert annotations [28].

- **4.3 Considerations for Dataset Suitability and Licensing:**

When selecting a dataset for training an LLM for commercial contract applications, several factors must be carefully considered. The size of the dataset is crucial, as larger datasets generally enable the model to learn more robust patterns. The relevance of the dataset to commercial contracts is paramount; a dataset primarily focused on other legal domains might not be as effective. If the goal is to train a model for specific tasks like clause identification or risk assessment, the presence and quality of annotations are essential. Finally, the licensing terms associated with the dataset must be thoroughly reviewed to ensure they permit the intended use, especially for commercial applications <sup>31</sup>. Obtaining large, high-quality, and publicly available datasets specifically for commercial contract generation can be challenging. In cases where public options are insufficient, exploring synthetic data generation techniques or curating proprietary datasets might be necessary <sup>32</sup>. It is also important to be aware of the legal implications of using different datasets, as concerns around copyright and the legal status of training data exist <sup>31</sup>. Developers must ensure compliance with licensing terms to avoid potential legal risks.

## **5. Pre-trained Large Language Models for Legal Applications:**

- **5.1 General Foundation Models and Their Applicability to Legal Tasks:**

A number of powerful general-purpose Large Language Models have demonstrated remarkable capabilities in understanding and generating human-like text. These models, such as OpenAI's GPT series (including the recently introduced GPT-4o <sup>5</sup>), Meta's Llama family <sup>18</sup>, Google's Gemini <sup>35</sup>, and Anthropic's models <sup>2</sup>, have been trained on vast amounts of diverse textual data. Their inherent ability to process and generate language makes them potentially suitable for various legal tasks, including summarizing lengthy legal documents, drafting initial versions of contracts, and answering legal queries <sup>1</sup>. However, to achieve optimal performance in the specialized domain of legal contracts, these general models typically require further adaptation through a process called fine-tuning <sup>34</sup>. This involves training the models on legal-specific datasets to enable them to better understand legal terminology, reasoning, and the unique



structures of legal documents. While general-purpose LLMs provide a strong foundation due to their broad language understanding, they need to be adapted to the specific nuances and terminology found in legal contracts.

- **5.2 Domain-Specific Legal LLMs: Capabilities and Training Data:**

In addition to general-purpose models, several LLMs have been specifically trained or fine-tuned on data from the legal domain. These domain-specific models often exhibit enhanced understanding of legal concepts and terminology right out of the box. **BloombergGPT** is one such model, trained on a substantial corpus of financial data, which also includes a significant amount of legal information<sup>37</sup>. While its primary focus is finance, its exposure to legal documents makes it potentially relevant for commercial contract analysis. **ChatLAW** is an open-source language model specifically trained using datasets in the Chinese legal domain<sup>17</sup>. While its training data and language focus differ from the user's requirements, it serves as a compelling example of a domain-specific legal LLM. **FinGPT** is another example, a lightweight language model pre-trained with financial data<sup>37</sup>, which, similar to BloombergGPT, might offer some transfer learning benefits for commercial contract analysis. **DeepSeek R1** is an open-source large language model that has shown potential in legal tech applications, including contract analysis and review<sup>58</sup>. The open-source nature of DeepSeek R1 is a significant advantage, providing greater transparency, opportunities for community contributions, and customization options. The potential advantage of using these domain-specific models lies in their pre-existing understanding of legal terminology and improved performance on legal tasks<sup>37</sup>. However, the availability of high-quality domain-specific LLMs tailored specifically for English commercial contracts is still an evolving area. Models trained on financial data might offer some initial understanding, and open-source models like DeepSeek R1 provide a flexible foundation for further fine-tuning.

- **5.3 Fine-tuning Strategies for Contract Analysis and Generation:**

Fine-tuning is a crucial process for adapting pre-trained LLMs to specific tasks within a particular domain<sup>34</sup>. This involves taking a model that has already learned general language patterns and further training it on a smaller, domain-specific dataset. Several fine-tuning techniques exist, including continued pre-training, where the model is trained on a large corpus of unlabeled domain-specific data, and supervised fine-tuning, where the model is trained on a smaller, labeled dataset relevant to the target task<sup>43</sup>. Parameter-efficient fine-tuning methods, such as Low-Rank Adaptation (LoRA), can be particularly useful for reducing the computational costs associated with fine-tuning large models<sup>34</sup>. For tasks like contract generation and risk identification, instruction tuning, which involves

training the model on specific prompts and desired outputs, can be highly effective<sup>50</sup>. The success of fine-tuning heavily relies on the quality and relevance of the training data<sup>33</sup>. The IBM Data Prep Kit is an open-source toolkit designed to streamline the often complex process of preparing unstructured data like legal contracts for fine-tuning<sup>46</sup>. It offers features for tasks such as converting PDFs to a structured format, redacting personally identifiable information (PII), and segmenting documents into manageable chunks<sup>46</sup>. Best practices for fine-tuning include clearly defining the desired outcome, selecting a suitable pre-trained model, carefully preparing and curating the training data, choosing appropriate hyperparameters, and rigorously evaluating the model's performance on a separate validation set<sup>34</sup>. Fine-tuning a pre-trained LLM on a dataset like CUAD, using tools like the IBM Data Prep Kit for data preparation, represents a viable strategy for developing a model specifically tailored for commercial contract analysis and risk evaluation.

## 6. Open-Source Resources for Building Legal LLM Applications:

- **6.1 Open-Source Contract Analysis Tools and Frameworks:**

Several open-source projects provide tools and frameworks that can be leveraged for building LLM-based applications for legal contract analysis. **contract-analyzer** is an open-source tool available on GitHub that utilizes the Meta-Llama-3-8B-Instruct model for contract analysis through a Retrieval-Augmented Generation (RAG) system<sup>63</sup>. This tool offers functionalities for processing PDF documents, performing semantic searches to retrieve relevant information, and generating responses using the integrated LLM<sup>63</sup>. It also allows for customization of parameters related to document chunking and information retrieval<sup>63</sup>. The existence of this project demonstrates a practical implementation of LLM-based contract analysis using a specific open-source LLM and provides a valuable starting point for developers interested in building similar applications. Its features, such as PDF processing and semantic search, directly align with the user's stated needs.

While not specifically focused on contract analysis, other open-source frameworks like LangChain and LlamaIndex (mentioned in<sup>33</sup>) provide powerful abstractions and tools for building various types of LLM-powered applications, including those for legal text processing.

- **6.2 Legal Text Processing Libraries (e.g., LexNLP, spaCy):**

A number of open-source Python libraries are available that offer functionalities for processing legal text, which can be essential for building the desired application.

**LexNLP** is an open-source Natural Language Processing (NLP) library specifically

designed for analyzing legal text <sup>64</sup>. It provides a wide range of features relevant to contract analysis, including tokenization (splitting text into words and sentences), named entity recognition (identifying key entities like dates, parties, and locations), and extraction of specific legal terms and financial figures <sup>64</sup>. Its focus on the legal domain makes it a valuable resource for preprocessing and extracting features from commercial contracts.

**spaCy** is another powerful open-source NLP library in Python known for its efficiency and industrial-strength capabilities <sup>70</sup>. It offers features like tokenization, part-of-speech tagging, dependency parsing (understanding the grammatical structure of sentences), named entity recognition, lemmatization (reducing words to their base form), and text classification <sup>70</sup>. Importantly, spaCy also provides support for integrating with Large Language Models through its spacy-llm package <sup>72</sup>. This integration allows developers to leverage the power of LLMs within spaCy's structured NLP pipelines for advanced legal text analysis tasks. spaCy's speed, efficiency, and extensibility make it a valuable tool for various legal text processing tasks, including those related to contract analysis.

**OpenNyAI** is an open-source NLP pipeline specifically designed for Indian legal documents <sup>77</sup>. While its focus on Indian law makes it less directly applicable to English commercial contracts, it demonstrates the availability of domain-specific open-source NLP tools. General-purpose NLP libraries like NLTK also offer a wide range of functionalities for text processing <sup>70</sup>.

- **6.3 Data Preparation Tools (e.g., IBM Data Prep Kit):**

As discussed earlier, the **IBM Data Prep Kit (DPK)** is an open-source toolkit that can significantly simplify the data preprocessing stage for LLM applications, including those in the legal domain <sup>46</sup>. Its features for handling various document formats (including PDFs), redacting sensitive information, deduplicating data, and chunking documents are particularly relevant for preparing legal contract data for training or use with LLMs. Utilizing such open-source data preparation tools can save considerable development time and effort.

## **7. Best Practices for Developing an LLM-Based Contract Application:**

- **7.1 Data Acquisition and Preprocessing:**

The foundation of an effective LLM-based contract application lies in the acquisition and meticulous preprocessing of a relevant and high-quality dataset of commercial contracts. This involves gathering data from various sources and employing data cleaning techniques to address inconsistencies, errors, and noise <sup>20</sup>. Standard preprocessing steps such as tokenization (breaking down text into smaller units), stemming or lemmatization (reducing words to their root form), and potentially anonymization through PII redaction (using tools like IBM DPK) are

crucial for preparing the data for model training<sup>20</sup>. Public contract databases like TermScout offer valuable insights into market standards and commonly used contract clauses through their extensive collection of analyzed contracts<sup>78</sup>. While direct download of the raw contract data for training might not be readily available, the structured data and analytics provided by such platforms can be invaluable for informing the development process and benchmarking the application's performance. Furthermore, data augmentation techniques can be employed to increase the size and diversity of the training data, potentially improving the model's generalization capabilities<sup>20</sup>.

- **7.2 Model Selection and Fine-tuning Techniques:**

Choosing the appropriate pre-trained LLM is a critical decision. Developers need to consider whether a general-purpose model fine-tuned on legal data or a domain-specific legal LLM would be more suitable for their specific requirements and available resources<sup>33</sup>. A common and often effective strategy is to start with a robust open-source general-purpose model, such as a Llama model or DeepSeek R1, and fine-tune it using a relevant legal contract dataset like CUAD<sup>34</sup>. To mitigate the computational demands of fine-tuning, exploring parameter-efficient methods like LoRA is recommended<sup>34</sup>. For the development of an application that can generate contracts and evaluate risks, instruction tuning, where the model is trained on specific instructions and examples for each task, is a crucial technique to consider<sup>50</sup>.

- **7.3 Evaluation Metrics and Validation Strategies:**

Thorough evaluation is essential to ensure the developed LLM-based contract application performs accurately and reliably. For contract generation, relevant evaluation metrics include fluency and coherence of the generated text, as well as its legal accuracy and adherence to relevant standards<sup>34</sup>. For risk evaluation, metrics such as precision, recall, and F1-score in identifying risky clauses are important<sup>34</sup>. It is crucial to use a separate validation dataset, distinct from the training data, to tune the model's hyperparameters and prevent overfitting, where the model learns the training data too well and performs poorly on unseen data<sup>22</sup>. Furthermore, involving legal professionals in the evaluation process to assess the practical utility and accuracy of the application is highly recommended<sup>1</sup>. Benchmarking the application's performance against existing commercial tools or open-source baselines, potentially using datasets like LegalBench, can provide a measure of progress and identify areas for improvement.

## **8. Conclusion and Recommendations**

The analysis indicates a significant potential for LLM-based applications to transform

the way legal professionals approach commercial contract generation and risk evaluation. The market is populated with a growing number of sophisticated commercial platforms offering a range of features designed to enhance efficiency and accuracy. Furthermore, a wealth of resources, including specialized datasets like CUAD, powerful pre-trained language models (both general and domain-specific), and valuable open-source tools and libraries, are available to support the development of such applications.

Based on the findings of this survey, the following recommendations are offered:

1. Begin by exploring the **CUAD dataset** for fine-tuning a suitable open-source pre-trained LLM, such as a model from the Llama family or DeepSeek R1. CUAD's focus on commercial contracts and expert annotations makes it an ideal resource for this task.
2. Leverage open-source libraries like **LexNLP** and **spaCy** for legal text processing and feature engineering. LexNLP's legal domain specificity and spaCy's efficiency and LLM integration capabilities offer complementary strengths.
3. Consider utilizing the **IBM Data Prep Kit** to streamline the data preprocessing pipeline. Its features for handling legal documents and preparing data for LLM training can save significant development time.
4. Establish a rigorous evaluation process that includes both automated metrics relevant to contract generation and risk evaluation, as well as thorough validation by legal professionals to ensure the practical utility and accuracy of the application.
5. Continuously monitor the evolving landscape of legal LLMs and open-source tools to identify new advancements and resources that can further enhance the application's capabilities.

By carefully considering these recommendations and leveraging the available resources, the development of an LLM-based application to assist lawyers in commercial contract generation and risk evaluation holds significant promise for improving efficiency, accuracy, and overall productivity within the legal profession.

## Works cited

1. How Large Language Models (LLMs) Can Transform Legal Industry - Custom AI Agents | Springs, accessed March 24, 2025, <https://springsapps.com/knowledge/how-large-language-models-llms-can-transform-legal-industry>
2. Robin AI | Review Contracts 80% Faster | Legal AI, accessed March 24, 2025, <https://www.robinai.com/>
3. Luminance - Legal-Grade™ AI, accessed March 24, 2025,

- <https://www.luminance.com/>
4. 9 Best Contract Analysis Software Tools to Speed Up Approvals - Percipient, accessed March 24, 2025, <https://percipient.co/best-contract-analysis-software/>
  5. The 9 best AI contract review software tools for 2025 | LEGALFLY, accessed March 24, 2025, <https://www.legalfly.com/post/9-best-ai-contract-review-software-tools-for-2025>
  6. AI-Powered Contract Management Solutions for Legal Teams, accessed March 24, 2025, <https://contractpodai.com/>
  7. Contract Management Artificial Intelligence | VISDOM AI, accessed March 24, 2025, <https://www.cobblestonesoftware.com/solutions/contract-management-ai>
  8. Accelerate Contract Reviews by 40% with Contract Risk Assessment ..., accessed March 24, 2025, <https://www.icertis.com/products/copilots/risk-assessment/>
  9. Using AI as a Contract Risk Assessment Tool - LexCheck Blog, accessed March 24, 2025, <https://blog.lexcheck.com/using-ai-as-a-contract-risk-assessment-tool-lc>
  10. Ivo - Leading AI Contract Review Software for Legal Teams, accessed March 24, 2025, <https://www.ivo.ai/>
  11. Contracts | Legaltech Hub, accessed March 24, 2025, <https://www.legaltechnologyhub.com/topics/contracts/>
  12. Can AI Write Legal Contracts? - Bloomberg Law, accessed March 24, 2025, <https://pro.bloomberglaw.com/insights/technology/can-ai-write-legal-contracts/>
  13. Business contract risk analysis with AI - CogniVis AI, accessed March 24, 2025, <https://cognivis.ai/examples/business-contract-risk-analysis-with-ai/>
  14. Hands-On Expert-Level Contract Summarization Using LLMs | Width.ai, accessed March 24, 2025, <https://www.width.ai/post/contract-summarization-using-llms>
  15. How to Use LLM to Extract Important Information from Legal Documents | by Addepto, accessed March 24, 2025, <https://medium.com/@addepto/how-to-use-llm-to-extract-important-information-from-legal-documents-a40ac6f974d8>
  16. How Legal Can Leverage Large Language Models to Improve Contract Negotiations, accessed March 24, 2025, <https://blog.lexcheck.com/how-legal-can-leverage-llms-improve-contract-negotiations>
  17. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases - SciSpace, accessed March 24, 2025, <https://scispace.com/pdf/chatlaw-open-source-legal-large-language-model-with-2p6m6i4m.pdf>
  18. ChatLaw: A Powerful LLM Tailored for Chinese Legal. 中文法律大模型 - GitHub, accessed March 24, 2025, <https://github.com/PKU-YuanGroup/ChatLaw>
  19. Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model - arXiv, accessed March 24, 2025, <https://arxiv.org/html/2306.16092v2>
  20. Open-Sourced Training Datasets for Large Language Models (LLMs) - Kili Technology, accessed March 24, 2025,



<https://kili-technology.com/large-language-models-llms/9-open-sourced-datasets-for-training-large-language-models>

21. Public Data Sets in the era of LLM's | by PATRICK O'CONNOR-READ - Medium, accessed March 24, 2025,  
<https://medium.com/@zolayola/public-data-sets-in-the-era-of-llms-0a4e89bda658>
22. nguha/legalbench · Datasets at Hugging Face, accessed March 24, 2025,  
<https://huggingface.co/datasets/nguha/legalbench>
23. legalbench/UsingLegalBench.ipynb at main · GitHub, accessed March 24, 2025,  
<https://github.com/HazyResearch/legalbench/blob/main/UsingLegalBench.ipynb>
24. Fine-tune ChatGPT-3.5 and GPT-4 with Weights & Biases - Colab, accessed March 24, 2025,  
[https://colab.research.google.com/github/wandb/examples/blob/master/colabs/openai/Fine\\_tune\\_OpenAI\\_with\\_Weights\\_and\\_Biases.ipynb](https://colab.research.google.com/github/wandb/examples/blob/master/colabs/openai/Fine_tune_OpenAI_with_Weights_and_Biases.ipynb)
25. HazyResearch/legalbench: An open science effort to benchmark legal reasoning in foundation models - GitHub, accessed March 24, 2025,  
<https://github.com/HazyResearch/legalbench>
26. theatticusproject/cuad · Datasets at Hugging Face, accessed March 24, 2025,  
<https://huggingface.co/datasets/theatticusproject/cuad>
27. theatticusproject/cuad-qa · Datasets at Hugging Face, accessed March 24, 2025,  
<https://huggingface.co/datasets/theatticusproject/cuad-qa>
28. Contract Understanding Atticus Dataset (CUAD) - Kaggle, accessed March 24, 2025,  
<https://www.kaggle.com/datasets/theatticusproject/atticus-open-contract-dataset-aok-beta/data>
29. CUAD - atticus-project, accessed March 24, 2025,  
<https://www.atticusproject.ai/org/cuad>
30. cuad\_unlimited-all-you-can-eat-license | LegalBench - Hazy Research, accessed March 24, 2025,  
[https://hazyresearch.stanford.edu/legalbench/tasks/cuad\\_unlimited-all-you-can-eat-license.html](https://hazyresearch.stanford.edu/legalbench/tasks/cuad_unlimited-all-you-can-eat-license.html)
31. Best Practices for Open Datasets for LLM Training Draft, accessed March 24, 2025,  
[https://assets.mofoprod.net/network/documents/Best\\_Practices\\_for\\_Open\\_Datasets\\_for\\_LLM\\_Training\\_Draft\\_-\\_Dataset\\_Convening\\_2024.pdf](https://assets.mofoprod.net/network/documents/Best_Practices_for_Open_Datasets_for_LLM_Training_Draft_-_Dataset_Convening_2024.pdf)
32. LLM training datasets - Glenn K. Lockwood, accessed March 24, 2025,  
<https://www.glennklockwood.com/garden/LLM-training-datasets>
33. How to Train LLM on Your Own Data in 8 Easy Steps - Airbyte, accessed March 24, 2025,  
<https://airbyte.com/data-engineering-resources/how-to-train-llm-with-your-own-data>
34. Fine-tuning Guide - Unsloth Documentation, accessed March 24, 2025,  
<https://docs.unsloth.ai/get-started/fine-tuning-guide>
35. Best Legal AI Tools for 2025 - Pocketlaw, accessed March 24, 2025,  
<https://pocketlaw.com/content-hub/legal-ai-tools>

36. Repository for organizing datasets and papers used in Open LLM. - GitHub, accessed March 24, 2025, <https://github.com/dsdanielpark/open-llm-datasets>
37. Building Domain-Specific LLMs: Examples and Techniques - Kili Technology, accessed March 24, 2025, <https://kili-technology.com/large-language-models-llms/building-domain-specific-llms-examples-and-techniques>
38. Large language models in legaltech: Demystifying fine-tuning - DraftWise, accessed March 24, 2025, <https://www.draftwise.com/blog/large-language-models-in-legaltech-demystifying-fine-tuning>
39. Domain Specific LLM: Special Applications of Large Language Models - IrisAgent, accessed March 24, 2025, <https://irisagent.com/blog/domain-specific-llm-revolutionizing-the-special-applications-of-large/>
40. ChatLaw-13B | AI Model Details - AIModels.fyi, accessed March 24, 2025, <https://www.aimodels.fyi/models/huggingFace/chatlaw-13b-pandavt>
41. How To Chose Perfect LLM For The Problem Statement Before Finetuning - Labellerr, accessed March 24, 2025, <https://www.labellerr.com/blog/how-to-choose-llm-to-suit-for-use-case/>
42. LLM Fine-Tuning: Use Cases, Best Practices, and Top 8 PEFT Methods - Kolena, accessed March 24, 2025, <https://www.kolena.com/guides/llm-fine-tuning-use-cases-best-practices-and-top-8-peft-methods/>
43. What is a Domain-Specific LLM? Examples and Benefits - Aisera, accessed March 24, 2025, <https://aisera.com/blog/domain-specific-llm/>
44. Large Language Models (LLMs): Challenges, Predictions, Tutorial - V7 Labs, accessed March 24, 2025, <https://www.v7labs.com/blog/large-language-models-llms>
45. LLM domain adaptation using continued pre-training — Part 1/4 | by Gili Nachum | Medium, accessed March 24, 2025, <https://medium.com/@gilinachum/llm-domain-adaptation-using-continued-pre-training-part-1-3-e3d10fcfdae1>
46. Fine-tuning LLMs for contract analysis using Data Prep Kit (DPK) - IBM Developer, accessed March 24, 2025, <https://developer.ibm.com/tutorials/dpk-fine-tuning-llms/>
47. A Beginner's Guide to Finetuning LLMs - Chatbase, accessed March 24, 2025, <https://www.chatbase.co/blog/finetuning-llms>
48. Pre-Training vs Fine Tuning: Choosing the Right Approach - Label Your Data, accessed March 24, 2025, <https://labelyourdata.com/articles/llm-fine-tuning/pre-training-vs-fine-tuning>
49. Fine-tuning LLMs | Thoughtworks United States, accessed March 24, 2025, <https://www.thoughtworks.com/en-us/insights/decoder/f/fine-tuning-llms>
50. LLM Fine Tuning Best Practices - Codoid, accessed March 24, 2025, <https://codoid.com/ai/llm-fine-tuning-best-practices/>
51. The Ultimate Guide to LLM Fine Tuning: Best Practices & Tools - Lakera AI,

- accessed March 24, 2025, <https://www.lakera.ai/blog/llm-fine-tuning-guide>
52. Fine-Tuning Large Language Models (LLMs): Techniques and Best Practices for Different Use Cases, accessed March 24, 2025, <https://sanchezsanchezsergio418.medium.com/fine-tuning-large-language-models-llms-techniques-and-best-practices-for-different-use-cases-97fab572a7a9>
  53. AI4Finance-Foundation/FinGPT: FinGPT: Open-Source Financial Large Language Models! Revolutionize We release the trained model on HuggingFace. - GitHub, accessed March 24, 2025, <https://github.com/AI4Finance-Foundation/FinGPT>
  54. FinGPT as a large FinLLM language model - Konfuzio, accessed March 24, 2025, <https://konfuzio.com/en/fingpt/>
  55. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases - ResearchGate, accessed March 24, 2025, [https://www.researchgate.net/publication/371943604\\_ChatLaw\\_Open-Source\\_Legal\\_Large\\_Language\\_Model\\_with\\_Integrated\\_External\\_Knowledge\\_Bases](https://www.researchgate.net/publication/371943604_ChatLaw_Open-Source_Legal_Large_Language_Model_with_Integrated_External_Knowledge_Bases)
  56. FinGPT/fingpt-sentiment-train · Datasets at Hugging Face, accessed March 24, 2025, <https://huggingface.co/datasets/FinGPT/fingpt-sentiment-train>
  57. Getting Started with FinGPT - Colab - Google, accessed March 24, 2025, [https://colab.research.google.com/github/AI4Finance-Foundation/FinGPT/blob/master/FinGPT\\_Training\\_LoRA\\_with\\_ChatGLM2\\_6B\\_for\\_Beginners.ipynb](https://colab.research.google.com/github/AI4Finance-Foundation/FinGPT/blob/master/FinGPT_Training_LoRA_with_ChatGLM2_6B_for_Beginners.ipynb)
  58. DeepSeek R1: Is This the Open-Source Legal Tech Breakthrough We've Been Waiting For?, accessed March 24, 2025, <https://www.geeklawblog.com/2025/01/deepseek-r1-is-this-the-open-source-legal-tech-breakthrough-weve-been-waiting-for.html>
  59. Data Prep Kit (DPK) - IBM Developer, accessed March 24, 2025, <https://developer.ibm.com/components/data-prep-kit>
  60. GitHub - data-prep-kit/data-prep-kit: Open source project for data ..., accessed March 24, 2025, <https://github.com/IBM/data-prep-kit>
  61. Preparing data for LLM Applications using Data Prep Kit - YouTube, accessed March 24, 2025, <https://www.youtube.com/watch?v=gxgHPFOqsNw>
  62. Unleash the potential of LLMs through the Data Prep Kit - IBM Developer, accessed March 24, 2025, <https://developer.ibm.com/blogs/awb-unleash-potential-llms-data-prep-kit/>
  63. ahmetkumass/contract-analyzer: Open-source tool for ... - GitHub, accessed March 24, 2025, <https://github.com/ahmetkumass/contract-analyzer>
  64. LexNLP - Open Source Legal, accessed March 24, 2025, <https://opensource.legal/projects/LexNLP>
  65. LexNLP - ContraxSuite, accessed March 24, 2025, <https://contraxsuite.com/lexnlp/>
  66. LexNLP Features - ContraxSuite, accessed March 24, 2025, <https://contraxsuite.com/lexnlp-features/>
  67. [1806.03688] LexNLP: Natural language processing and information extraction for legal and regulatory texts - arXiv, accessed March 24, 2025, <https://arxiv.org/html/1806.03688>
  68. About LexNLP — LexNLP 2.3.0 documentation - Read the Docs, accessed March 24, 2025, <https://lexpredict-lexnlp.readthedocs.io/en/latest/about.html>
  69. Fine-grained Contract NER using instruction based model - ACL Anthology,

- accessed March 24, 2025, <https://aclanthology.org/2023.paclic-1.89.pdf>
70. NLP Libraries in Python - GeeksforGeeks, accessed March 24, 2025, <https://www.geeksforgeeks.org/nlp-libraries-in-python/>
  71. Top 8 Python Libraries For Natural Language Processing (NLP) in 2025 - Analytics Vidhya, accessed March 24, 2025, <https://www.analyticsvidhya.com/blog/2021/05/top-python-libraries-for-natural-language-processing-nlp-in/>
  72. spaCy · Industrial-strength Natural Language Processing in Python, accessed March 24, 2025, <https://spacy.io/>
  73. Linguistic Features · spaCy Usage Documentation, accessed March 24, 2025, <https://spacy.io/usage/linguistic-features>
  74. spaCy 101: Everything you need to know, accessed March 24, 2025, <https://spacy.io/usage/spacy-101>
  75. Using spaCy for natural language processing (NLP) in Python - Domino Data Lab, accessed March 24, 2025, <https://domino.ai/blog/natural-language-in-python-using-spacy>
  76. Language Processing Pipelines · spaCy Usage Documentation, accessed March 24, 2025, <https://spacy.io/usage/processing-pipelines>
  77. OpenNnAI : An efficient NLP Pipeline for Indian Legal documents - GitHub, accessed March 24, 2025, <https://github.com/OpenNnAI/OpenNnAI>
  78. What is the Public Contract Database? - Blog, accessed March 24, 2025, <https://blog.term Scout.com/what-is-the-public-contract-database>
  79. Public Contract Database - Term Scout, accessed March 24, 2025, <https://www.term Scout.com/public-contract-database>
  80. Term Scout - Get Access | Soverin, accessed March 24, 2025, <https://soverin.ai/product/term-scout/>
  81. Terms & Conditions - TermScout, accessed March 24, 2025, <https://www.term Scout.com/terms-conditions>
  82. TermScout Reviews - Read Customer Reviews of Term Scout.com, accessed March 24, 2025, <https://term Scout.tenere team.com/>
  83. contract-certification - TermScout, accessed March 24, 2025, <https://www.term Scout.com/contract-certification-0-0-0-0>
  84. Contract API Solutions - TermScout, accessed March 24, 2025, <https://www.term Scout.com/contract-api-solutions>
  85. TermScout, accessed March 24, 2025, <https://www.term Scout.com/>
  86. Blog | TermScout, accessed March 24, 2025, <https://blog.term Scout.com/>