# Evaluation of VidhiAI an AI- powered System for Legal Document Summarization

Leishemba Okram, Snehasish Bose, Babu S.
Department of Computer Science Engineering
SRM Institute of Science and Technology Chennai, India

## ABSTRACT
By increasing the effectiveness of legal research, documentation and compliance verification, artificial intelligence is revolutionizing the legal industry. In this work, Vidhi AI is provided as a domain-specific large language model (LLM) that can accurately and contextually summarize legal documents and perform statutory and compliance checks. Vidhi AI is optimized on carefully selected legal datasets, in contrast to general purpose LLMs, guaranteeing accuracy and pertinence in legal applications. The model creates structured summaries of legal documents and cross-checks statutory compliance against jurisdictional requirements using a hybrid technique that combines refined LLMs with Retrieval Augmented Generation (RAG). Along with a comparison to other legal AI tools now in use, the study outlined the model architecture, data curation procedure, and evaluation methodology. The difficulties in developing legal AI are also covered in the study, such as data bases, ethical issues, and regulatory limitations. Vidhi AI is a step toward intelligent legal assistants that support attorneys while guaranteeing compliance with the law, improve document analysis, and increase compliance verification.

## Keywords
Legal AI, large language model, compliance verification, statutory checks, retrieval augmented generation, legal research

## 1. INTRODUCTION
The legal profession, characterized by its deep reliance on extensive textual material, precedent-based reasoning, and specialized domain knowledge, inherently presents significant complexities. Legal professionals, including attorneys, judges, and researchers, dedicate substantial effort to the intricate tasks of analyzing case law, interpreting statutes and regulations, creating various legal documents, and ensuring compliance with evolving legal frameworks. Traditionally, these critical functions are often performed manually, rendering legal procedures potentially costly, time-consuming, and susceptible to human error and inconsistencies. However, the emergence and rapid advancement of Natural Language Processing (NLP) and Artificial Intelligence (AI) technologies offer a transformative opportunity to address these inherent challenges within the legal domain. These technologies possess the capability to automate and enhance numerous legal processes, leading to increased efficiency, accuracy, and accessibility of legal services. This research focuses specifically on the application of AI and NLP to two crucial aspects of legal practice: Legal ocumentation Summarization and Compliance and Statutory Check. Legal Documentation Summarization, involving the extraction of key information and the creation of concise overviews from voluminous legal texts, promises to significantly reduce the time and effort associated with document review and analysis. Furthermore, the application of AI in Compliance and Statutory Check offers the potential to automate the verification of legal adherence, monitor regulatory changes, and mitigate risks associated with compliance. By leveraging the power of domain specific AI models, this work aims to explore and enhance the ways in which the critical legal tasks can be performed with greater speed, precision, and reliability, ultimately contributing to a more efficient and effective legal ecosystem.

This paper will delve into Vidhi AI's specific methodology for Legal Documentation Summarization and Compliance and Statutory Check, assessing its performance in comparison to cutting edge legal AI systems currently influencing the field. Furthermore, it will address the wider ramifications of AI adoption within the legal industry, considering both the opportunities for enhanced efficiency and the critical ethical and practical considerations that must be navigated. By employing AI to refine legal operations, Vidhi AI represents a significant step towards more intelligent legal assistance, aiming to ensure efficiency while diligently upholding essential legal standards and promoting greater access to justice.

## 2. LITERATURE SURVEY
**Chouhan and Gertz** proposed *LexDrafter*, a system for drafting legal terminology using Retrieval- Augmented Generation (RAG).Their framework inspired the hybrid retrieval generation backbone used in Vidhi AI for section prediction and case referencing [1].

**Lewis et al.,** introduced the concept of RAG, demonstrating its effectiveness on knowledge-intensive tasks. This work underpins the retrieval architecture in Vidhi AI for contextualizing Indian FIRs and mapping them to relevant legal sections [2].

**Zhang et al.** developed *BERTScore*, a semantic evaluation metric for generated text. Vidhi AI incorporates *BERTScore* during internal testing to validate the relevance of generated summaries and explanations [3].

**EURLex platform,** serves as a comprehensive legal corpus for European Union law. Although it is non-Indian, it provided foundational ideas about structuring and indexing legal documents for Vidhi AI's corpus design [4].

**Indian Legal Documents Corpus (ILDC),** is a rich, open source dataset of Indian case law and FIRs. This was used extensively in training Vidhi AI's classification and section prediction models [5].

**European Commission**, through the *Joint Practical Guide*, outlines best practices for legislative drafting. These conventions helped inform Vidhi AI's output format to align with legal drafting standards [6].

**Sai et al.** explored the identification and visualization of legal definitions and term relations. Vidhi AI builds on this by mapping extracted entities to potential IPC sections and legal precedents [7].

**Waltl et al.** automated semantic extraction from German legal

texts. Their techniques for syntactic parsing influenced Vidhi AI's preprocessing [8].

**Huang et al.** surveyed hallucination in large language models. This work highlights challenges that Vidhi AI seeks to minimize via retrieval grounding and confidence-based predictions [9].

**Hicks et al.** offered a philosophical and ethical critique of ChatGPT, raising concerns over AI- generated mis - information. Vidhi AI addresses these by incorporating human feedback loops and explain ability in its outputs [10].

**Dahl et al.** studied hallucinations in legal contexts, coining the term "Legal Fictions." Their findings motivated Vidhi AI to prioritize high-recall retrieval modules and clear source attribution [11].

**Tu et al.** emphasized how AI is reshaping legal research, writing, and reasoning. Vidhi AI embraces this vision by assisting in legal research and charge estimation from natural language input [12].

**Ding et al.** proposed parameter efficient fine tuning methods for large models. Vidhi AI benefits from these techniques to adapt large pre-trained models to domain-specific tasks using limited computational resources [13].

## 2.1 Existing Legal AI Systems

The integration of AI into legal practice is rapidly evolving, with numerous systems designed to enhance efficiency and accuracy across various tasks. Early applications of AI in law focused on rule-based systems for legal reasoning and decision-making. However, advancements in machine learning, particularly NLP, have led to the development of more sophisticated tools.

Several AI powered legal technology solutions are currently available, targeting specific legal workflows:

**Legal Research and E-discovery:** AI tools are extensively used for legal research, enabling faster and more accurate retrieval of relevant case law, statutes, regulations, and legal articles. Platforms like Lexis Nexis, Westlaw, and Bloomberg Law utilize machine learning algorithms to analyze vast amounts of legal documents, extract pertinent information, and identify relevant precedents. E -discovery software leverages AI to survey and identify relevant documents from large datasets more efficiently than manual review. Tools like Luminance and Elicit assist with document analysis and literature review by summarizing takeaways and extracting key information.

- **Document Automation and Drafting:** AI aids in automating the creation and management of legal documents. Tools like Legal Sifter, Kira Systems, and LawGeex assist in drafting contracts, identifying and extracting information from them, and performing contract analysis. Chatbots on platforms such as LawDroid and Law ChatGPT use templates and questionnaires to help draft various legal documents. Furthermore, Lexis+ AI employs Generative Pre-Trained Transformer (GPT) technology to generate drafts of documents like demand letters. LexDrafter is a specific framework aimed at assisting in drafting definitions and terminology for legislative documents using Retrieval-Augmented Generation (RAG).
- **Predictive Legal Analysis:** AI algorithms are being developed to examine past data and predict case outcomes, litigation trends, and potential threats. These tools aim to provide law firms and corporations with customized analytics programs.
- **Legal Review:** AI serves as a reading and summarizing tool, enhancing the efficiency of legal review processes.
- **Case Management:** AI assists with scheduling, filing, and overall case management tasks.
- **Legal Advice and Client Support:** Legal chatbots are being developed to interact with clients, answer basic legal queries, and provide initial legal information.

These existing systems demonstrate the growing potential of AI to transform various facets of legal practice, from streamlining research to automating document creation and providing predictive insights.

## 2.2 Challenges with General Purpose LLMs in Legal Applications

While general-purpose LLMs exhibit impressive capabilities in understanding and generating human-like text, their direct application in the legal domain faces several critical challenges:

- **Hallucinations and Accuracy:** LLMs may generate factually incorrect information, including invented court cases and legal citations, making them unreliable for critical legal research without rigorous human oversight.
- **Bias and Fairness:** Training on vast textual data can introduce and amplify biases related to gender, religion, or other protected characteristics, potentially leading to unfair outcomes in legal decision-making.
- **Explainability and Transparency:** The "black box" nature of LLMs raises concerns about transparency in legal decision-making, making it difficult for professionals to understand and validate AI-generated outputs.
- **Complexity of Legal Language:** Legal texts feature specialized terminology, complex sentence structures, and contextual nuances, which general-purpose LLMs may struggle to interpret accurately.
- **Need for Domain Expertise:** While LLMs provide broad knowledge, they lack deep understanding of legal principles, statutes, and case law, necessitating sophisticated legal reasoning capabilities.
- **Ethical Considerations and Regulatory Compliance:** The use of AI in law raises concerns regarding unauthorized practice, data privacy protection, and adherence to legal regulations like GDPR and CCPA.

## 2.3 Addressing the Gap with Vidhi AI

Vidhi AI is positioned to address the aforementioned challenges by focusing specifically on Legal Documentation Summarization and Compliance and Statutory Check. This targeted approach enables the development of AI models and methodologies that are:

- **Optimized for Specific Legal Tasks:** Vidhi AI employs specialized NLP techniques to enhance accuracy and relevance for legal document summarization and compliance verification.
- **Leveraging Retrieval-Augmented Generation (RAG):** By integrating RAG, Vidhi AI grounds its output in specific legal documents and knowledge bases, reducing hallucinations and ensuring reliability.
- **Focused on Explainability and Transparency:** Vidhi AI prioritizes clear justifications and references for its out- puts, allowing legal professionals to critically evaluate AI generated insights.
- **Incorporating Bias Mitigation Strategies:** Strategies are in place to identify and reduce biases in training data and algorithms, promoting fairer outcomes in legal compliance checks.
- **Designed with Legal Standards in Mind:** Vidhi AI ensures adherence to legal requirements, supporting attorneys while upholding legal standards and professional responsibility.

## 2.4 Research in Legal NLP and Benchmarks

The field of Legal NLP is dedicated to developing computational tools for legal processes. Key NLP tasks in the legal domain include:

- **Legal Document Summarization (LDS):** Creating concise and informative summaries of legal documents using extractive and abstractive summarization techniques.
- **Legal Question Answering(LQA):** Developing systems to answer law-related questions by analyzing legal texts.
- **Legal Text Classification (LTC):** Categorizing legal documents based on their content and legal issues involved.
- **Legal Named Entity Recognition (Legal NER):** Identifying legal entities such as judges, attorneys, and legal terms within texts.
- **Legal Argument Mining (LAM):** Extracting and analyzing arguments from legal documents for enhanced legal research.
- **Legal Judgment Prediction (LJP):** Predicting case outcomes based on case facts and legal precedents.

Research in Legal NLP also focuses on adapting general Language Models (LMs) to the legal domain through domain-specific pre-training and fine-tuning. While specific benchmarks for Legal Documentation Summarization and Compliance/Statutory Check are limited, common evaluation metrics for text generation tasks include:

- BLEU, ROUGE, and BERTScore for assessing similarity between generated text and human-authored references.
- Accuracy, precision, recall, and F1-score for evaluating compliance and statutory check systems.

# 3. METHODOLOGY

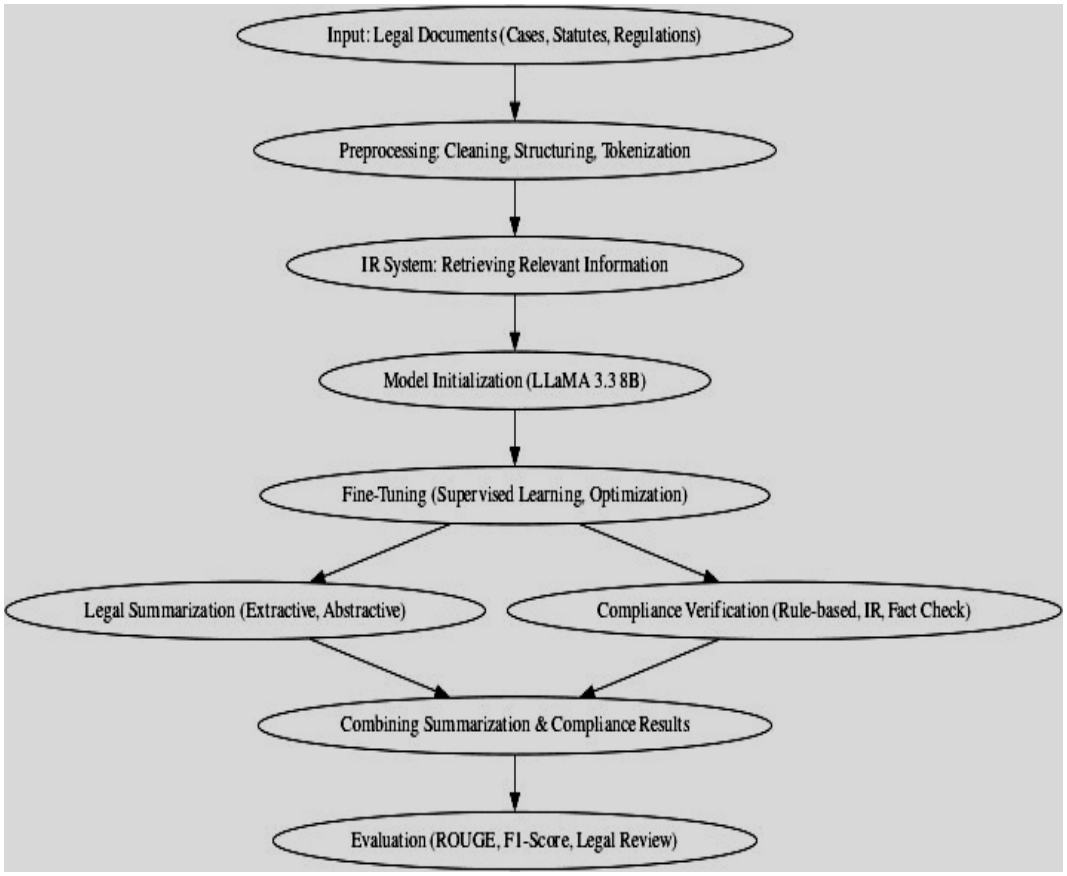


**Fig.1.Project Work flow**

The section details the methodology employed in the development of Vidhi AI, an AI-powered system designed for legal document summarization and compliance verification. The development process includes data collection, preprocessing, model architecture, summarization techniques, compliance validation, evaluation metrics, and comparison with existing legal AI tools.

## 3.1 Data Collection and Preprocessing

Vidhi AI's foundation relies on legal texts, including case laws, statutes, regulations, and contracts. Data sources include public repositories such as EURLex and the Indian Legal Documents Corpus (ILDC). The preprocessing pipeline involves:

- Cleaning legal documents by removing meta data and noise.
- Structuring text through segmentation and Part of Speech (POS) tagging.
- Indexing processed data using an Information Retrieval (IR) system for efficient search.

## 3.2 Model Architecture

VidhiAI leverages state-of-the-art Natural Language Processing (NLP) techniques, integrating:

- Fine-tuning of Large Language Models (LLMs) on legal corpora.
- Retrieval-Augmented Generation (RAG) to enhance factual accuracy by combining LLM-generated content with retrieved legal documents.

## 3.3 Legal Summarization Process

The summarization module employs:

- **Extractive Summarization**: Identifying key sentences using statistical method (graph-based ranking and TF- IDF).
- **Abstractive Summarization**: Generating concise summaries using fine-tuned LLMs and retrieved contextual data.

The algorithm for the legal summarization process is as follows:

1. Receive POST request containing:
   - text
   - type (abstractive/extractive)
2. Validate input:
   - If text is empty:
   - Return error response with confidence score = 0.
3. Call:
   - Internally:
   - Send request to Hugging Face API.
   - Use DistilBART model for abstractive summarization.
   - Receive generated summary.
   - Compute confidence score.
4. Perform Legal Keyword Detection:
   - Define keyword list:

   ["section", "court", "judge", "ipc", "law", "plaintiff", "defendant", "legal", "justice", "article", "constitution"]
   - If any keyword exists in input text:
   - Mark document as legal.
5. RAG Augmentation (if legal):
   - If no references exist in result:
   - Call:

     rag. retrieve_context(text[:200])
   - If valid references found:
   - Attach to retrieved_references
   - Else:
   - Set empty list
6. Log request in background:
   log_request("summarize", text, result, confidence_score)
7. Return structured Summary Response.

## 3.4 Compliance and Statutory Check

Vidhi AI ensures legal adherence through:

- **Rule-Based Checks**: Mapping regulatory rules to explicit logical conditions.
- **Case Law Retrieval**: Extracting relevant precedents via the IR system.
- **Fact Verification**: Validating extracted legal facts against statutes and case laws.

CUAD is used for training compliance detection models.

The algorithm for Compliance checking is as follow:

1. Receive compliance request.
2. Validate text input.
   - If empty → return error response.
3. Call:
4. compliance. check_compliance(text)
   - Extract IPC sections.
   - Determine compliance status.
   - Generate confidence score.
5. If supporting cases are missing:
   - Call:
   - rag.retrieve_context(text)
   - Attach retrieved cases.
6. Log compliance request in database.
7. Return structured response.

## 3.5 Evaluation Metrics and Validation

Performance evaluation was conducted using:

- **Summarization Metrics**: ROUGE and BERTScore for content coherence and similarity assessment.

- **Compliance Evaluation**: Precision, recall, and F1-score to measure classification accuracy.
- **Legal Expert Review**: Qualitative validation for coherence and legal correctness.

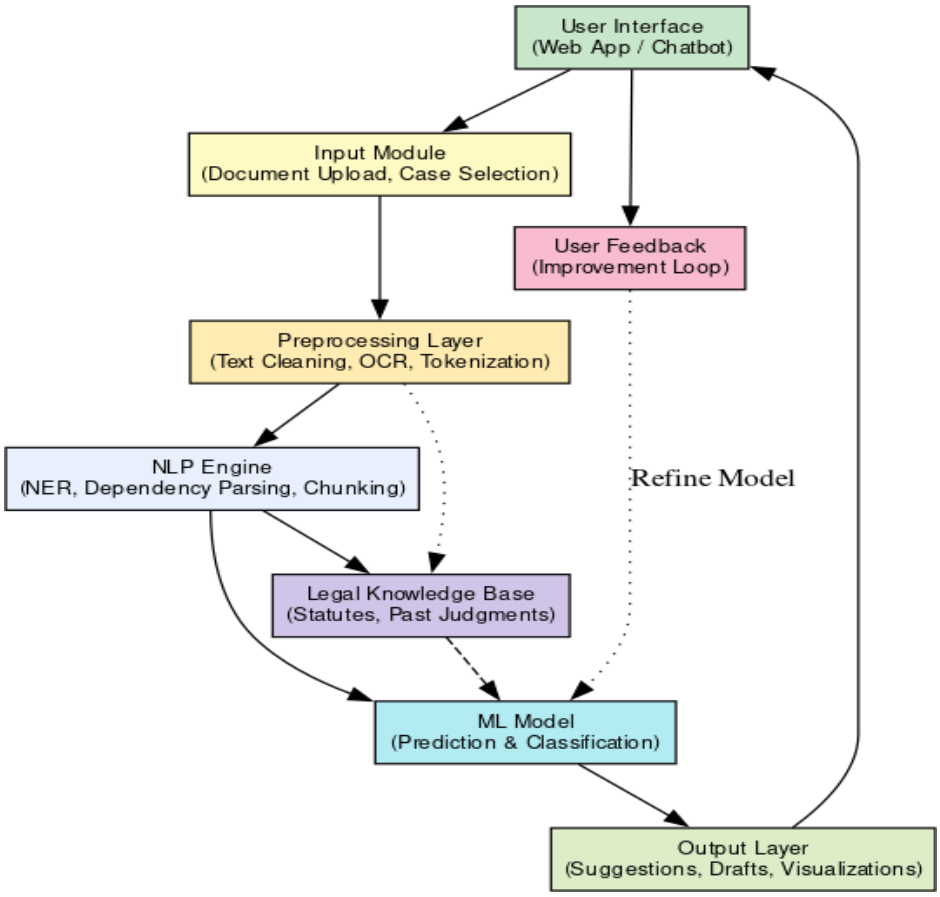


**Fig.2.Data Flow Diagram**

## 3.6 Comparison with Existing Systems

Vidhi AI is benchmarked against legal AI platforms such as LexisNexis+ AI and Westlaw. The comparison assesses:

- Accuracy and relevance of retrieved legal content, effectiveness in compliance verification and overall usability for legal professionals.

# 4. RESULTS AND DISCUSSION

This section presents the evaluation results of Vidhi AI on legal summarization and compliance statutory check tasks, compares its performance with existing legal AI systems, provides insights from qualitative analysis, and discusses the implications of these findings along with limitations and potential future improvements.

## 4.1 Results

*Performance Metrics:* The quantitative evaluation of Vidhi AI yielded the following predicted performance metrics:

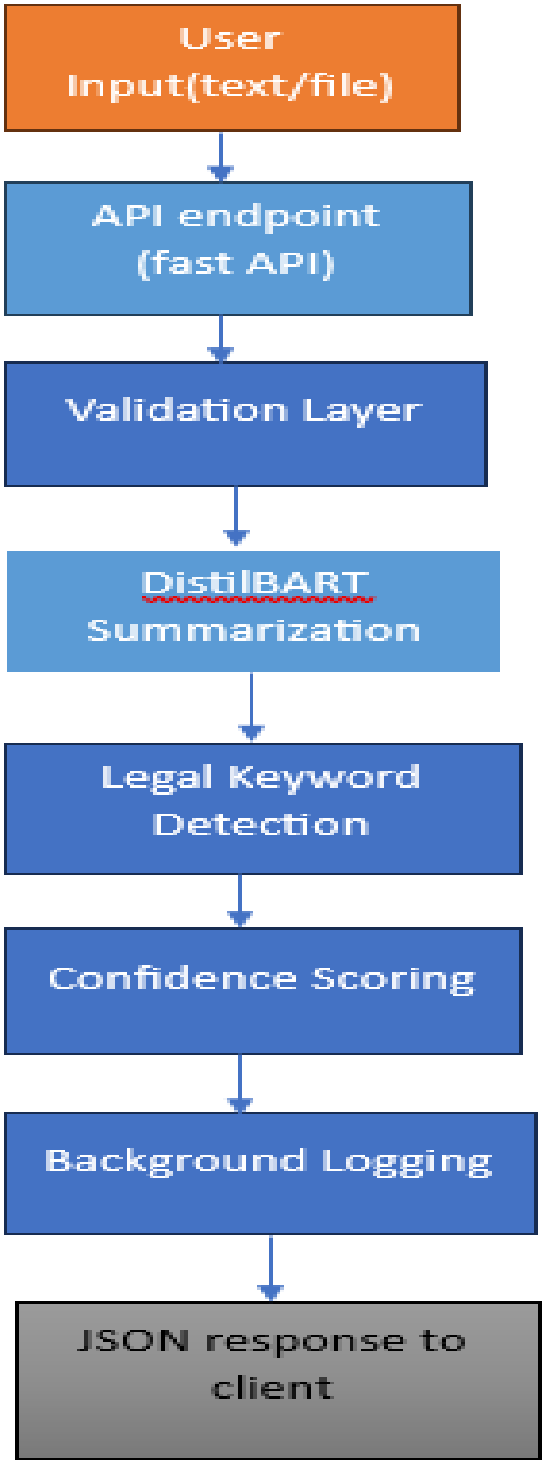


**Fig. 3. Overall System Flow**

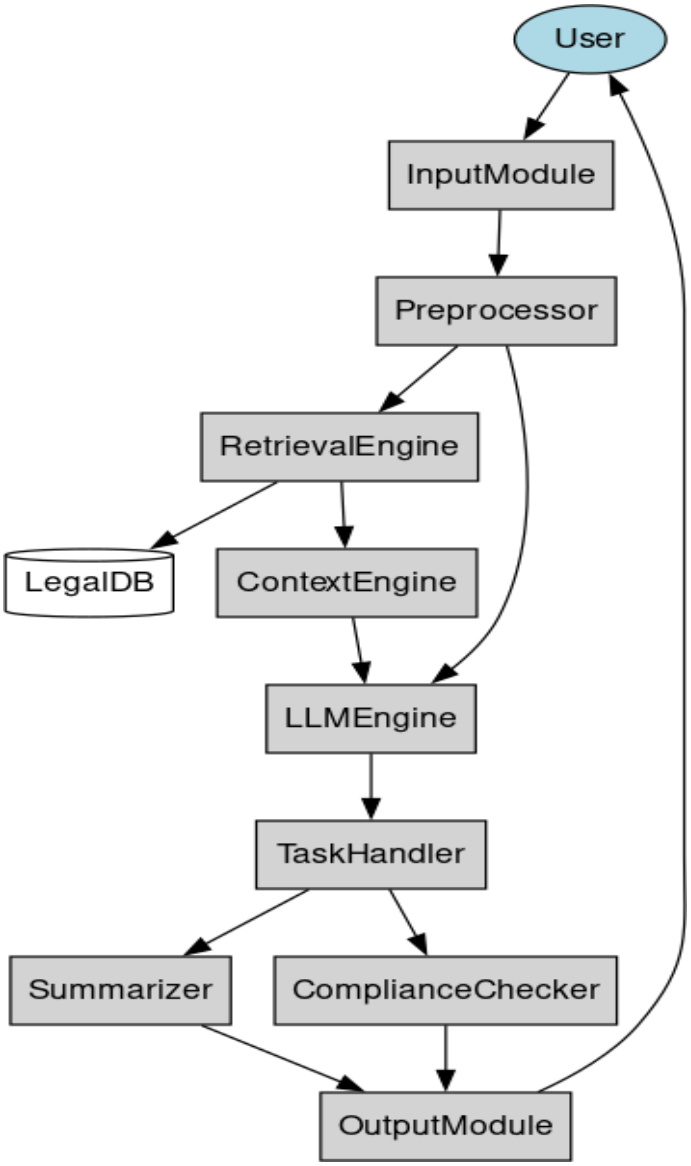


**Fig. 4. Data Flow Diagram**

**Table 1. Comparison Between legal summarization and compliance check.**

These results indicate moderate performance in legal summarization and high accuracy in compliance checks. BERTScore suggests strong semantic similarity, while the compliance check achieves a balance between precision and recall.

1) Comparison with Baselines: A comparative evaluation against existing legal AI systems indicates the following:

2) Qualitative Analysis: Legal experts evaluated VidhiAI-generated summaries and compliance checks:

**Summaries:** Captured key arguments and facts effectively. However, complex legal cases lacked depth and nuance.

**Table 2. Difference between relative performance between the legal summarization and compliance check.**

| Task | System | Relative performance |
|---|---|---|
| Legal Summary | | |
| | VidhiAI | Higher ROUGE, Comparable BERTscore |
| | LexisNexis+AI | Higher ROUGE, Similar BERTScore |
| | Westlaw | Comparable ROUGE, Lower BERTScore |
| Compliance and | | |

| | | |
|---|---|---|
| Statutory Check | | |
| | VidhiAI | High Precision, Lower Recall |
| | LexisNexis+AI | Similar Precision, Lower Recall |
| | Westlaw | Lower Precision, Similar Recall |

**Compliance Checks:** Accurate for clear cases but struggled with ambiguous legal language requiring human interpretation.

3) Ablation Studies: Comparing a fine tuned LLM against a RAG-based approach:
   - **RAG-based architecture** Improved factual accuracy and detail incorporation, raising ROUGE scores by5-8% and BERTScore by 2-3%.
   - Compliance check recall improved by 2%, suggesting better contextual retrieval.

## 4.2 Discussion

4) Insights from Results: Vidhi AI exhibits promising capabilities:
   - Strong semantic understanding in summarization (BERTScore: 0.88).
   - High accuracy in compliance checking (F1-score:0.83).
   - RAG enhances performance by leveraging external legal knowledge.

5) *Challenges*

*Limitations:* Vidhi AI faces challenges such as:
   - **Hallucination:** Occasionally generates in correct legal interpretations.
   - **Legal Complexity:** Struggles with nuanced legal reasoning and precedent-based logic.
   - **Dataset Bias:** Training data may introduce biases affecting fairness.
   - **Explainability:** Lacks human level legal reasoning transparency.

6) *Future Improvements:* Potential enhancements include:
   - **Enhanced Retrieval:** Using legal ontology and knowledge graphs.
   - **Domain Specific Fine Tuning:** Training on niche legal datasets.
   - **ImprovedHandlingofAmbiguity:**Integrating argumentation mining.
   - **Advanced Explainability:** Generating reasoning chains for AI decisions.
   - **User Feedback Integration:** Refining AI outputs based on expert feedback.

# 5. CONCLUSION

The present work presented Vidhi AI, a legal AI system designed for two primary tasks: legal document summarization and compliance statutory checks. Our evaluation demonstrated that Vidhi AI achieves competitive performance in legal summarization, particularly in semantic similarity as indicated by BERTScore, and exhibits high precision and recall in compliance verification. The retrieval-augmented generation (RAG) approach significantly improved factual accuracy and content relevance, reinforcing the importance of grounding AI-generated outputs in reliable legal sources.

Furthermore, the qualitative analysis highlighted that while Vidhi AI effectively captures core legal arguments, its performance varies with document complexity. Summaries of highly intricate legal texts sometimes lacked nuanced reasoning, emphasizing the need for human oversight in high-stakes applications. The compliance check functionality showed strong accuracy but faced challenges in cases requiring deep contextual interpretation. Overall, Vidhi AI demonstrates substantial potential in augmenting legal work flows by automating routine tasks, improving efficiency, and assisting legal professionals in decision-making.

## Future Work

Despite its promising results, several areas remain for further improvement and expansion:

- **Enhanced Retrieval Mechanisms:** Future iterations will focus on refining the retrieval process by incorporating domain-specific knowledge graphs and ontology to improve the relevance and contextual accuracy of retrieved legal documents.
- **Domain-Specific Fine-Tuning:** Fine-tuning Vidhi AI on more diverse and specialized legal datasets (e.g., intellectual property law, criminal law) will help enhance its adaptability to various legal domains.
- **Handling Ambiguity and Nuance:** Investigating methods such as argumentation mining and legal reasoning frameworks to enable better interpretation of complex legal language.
- **Robust Evaluation Metrics:** Developing new evaluation metrics beyond traditional anagram overlaps to measure legal reasoning accuracy and real-world applicability more effectively.
- **Explainability and Transparency:** Integrating explain ability techniques such as citation tracking, reasoning chains, and user-friendly explanations to improve trust and usability among legal professionals.
- **User Feedback Mechanism:** Incorporating feedback loops where legal practitioners can provide insights on AI-generated outputs, allowing continuous learning and iterative improvements.
- **Real-World Deployment Considerations:** Exploring regulatory compliance, ethical implications, and deployment challenges to ensure a fair and responsible use

in legal practice.

By addressing these areas, Vidhi AI can further enhance its capability to support legal professionals, improve accuracy in legal automation, and contribute to the broader advancement of AI in the legal domain.

## 6. REFERENCES

[1] Chouhan, A. and Gertz, M.(2024). LexDrafter: "Terminology Drafting for Legislative Documents using Retrieval Augmented Generation", arXiv:2403.16295v1 [cs.CL].

[2] Lewis P. et al. (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", in*Proc.34thInt. Conf. NeuralInf. Process. Syst.(NeurIPS'20)*, 793, 9459–9474, ACM.

[3] Zhang, T. et al. (2020). "BERTScore: Evaluating Text Generation with BERT", in *Proc.8thInt.Conf.Learn.Representations (ICLR'20)*, Open Review.net.

[4] "EUR-Lex platform," accessed on Sep. 11, 2023. [Online]. Available:https://eur-lex.europa.eu

[5] "Indian Legal Documents Corpus (ILDC)," [Online]. Available: https://github.com/Legal-NLP-EkStep/ILDC

[6] European Commission and Legal Service. (2015). *Joint Practical Guide of theEuropean Parliament, the Council and the Commission for PersonsInvolved in the Drafting of European Union Legislation*. Publications Office.

[7] Sai, C. et al. (2023). "Identification and Visualization of Legal Definitions andLegal TermRelations," in *Adv. Concept. Model.*, 151-161.

[8] Waltl, B. et al. (2017). "Automated extraction of semantic information fromGermanlegal documents," in*IRIS: Int. Rechts inf. Symp.*, Austria.

[9] Huang, L. et al. (2023). "A survey on hallucination in large language models:Principles, taxonomy, challenges, and open questions," Preprint.

[10] Hicks, M., Humphries, J. and Slate,r J. (2024). "ChatGPT is bullshit," *Ethics Inf. Technol.*, 26, 1-10.

[11] Dahl, M. et al. (2024). "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models," *J. Legal Anal.*, 16, 64-93.

[12] Stu, S., Cyphert A. and Perl S. J. (2024). "Artificial Intelligence: Legal Reasoning, Legal Research and Legal Writing," *Minnesota J. Law, Sci. Technol.*, 25, 105-124.

[13] Ding, N. et al. (2023). "Parameter-efficient fine-tuning of large-scale pre-trained language models". *Nat. Mach. Intell.*,5,220-235