**Intelligent Legal Framework**

Aaditya Bhargav, Devansh Garg, Mukul Vyas, Pushkar Maurya, Rahul Kishore Gorai

Department of Computer Science and Engineering, IIIT Delhi

**ABSTRACT**

*This paper presents an innovative Intelligent Legal Framework (ILF) that integrates various AI-driven tools and functionalities into a unified platform. The proposed solution aims to address existing challenges in legal research and practice, including inefficient legal research, limited predictive capabilities, language barriers, and fragmented legal services. The ILF comprises components such as Legal Assistant and Drafter, Legal Summarizer, Legal Chatbot, and Legal Multi-PDF Chat, leveraging advanced AI techniques like predictive analytics, natural language processing, and multi-modal analysis. The platform is designed to enhance productivity, improve decision-making, and increase access to justice through affordable and accessible legal services.*

**KEY WORDS**

*Intelligent Legal Framework, AI-driven platform, legal research, predictive analytics, natural language processing, multi-modal analysis, legal services, access to justice****.***

**1 INTRODUCTION**

The legal profession has traditionally been characterized by rigorous research, meticulous analysis, and a reliance on established precedents. However, the rapid advancements in artificial intelligence (AI) and machine learning (ML) technologies have ushered in a new era of transformation, promising to streamline legal processes and enhance the overall efficiency of legal services. Despite these technological advancements, several challenges continue to persist within the legal domain, hindering its full potential. Conventional legal research methods are often time-consuming and lack comprehensive analysis capabilities. Legal professionals must sift through vast repositories of case laws, statutes, and legal documents, a process that can be tedious and prone to oversights or missed connections. The sheer volume of information and the complexity of legal reasoning demand more efficient and automated approaches to legal research.

Existing legal tools and resources lack sophisticated predictive analysis features, preventing legal professionals from accurately anticipating case outcomes and identifying emerging trends. The ability to forecast legal outcomes and understand the implications of various scenarios is crucial for strategic decision-making and effective legal counsel. The global nature of legal practice necessitates the ability to comprehend and communicate legal matters across multiple languages. However, limited multilingual capabilities in current legal tools and resources hinder the seamless understanding and translation of legal documents, posing significant challenges in cross-border legal proceedings and international collaborations.

Legal services are often fragmented, with specialized tasks and functions segregated into separate tools or platforms. This fragmentation can impede the effective utilization of specialized skills and expertise, leading to inefficiencies and potential oversights in legal processes.

In response to these challenges, this project introduces the Intelligent Legal Framework (ILF), an innovative and integrated AI-driven platform designed to revolutionize legal research and practice. The ILF aims to address the existing gaps by leveraging cutting-edge AI and ML technologies, enabling legal professionals to work more efficiently, make informed decisions, and provide accessible legal services to a broader population.

**2 PROBLEM STATEMENT**

The legal profession faces several challenges, including inefficient legal research processes, limited predictive capabilities, language barriers, and fragmented legal services. Current legal research methods are often time-consuming and lack comprehensive analysis, hindering legal professionals' ability to anticipate case outcomes and trends effectively. Moreover, existing legal tools lack sophisticated predictive analysis features, further exacerbating this problem. Language barriers and limited multi-modal capabilities also pose significant obstacles to comprehensive document understanding, hampering effective collaboration and communication. Finally, legal services are typically separated, making it difficult to leverage specialized skills and tools in a cohesive manner. Our research aims to address these issues by developing an integrated platform that streamlines legal research, provides predictive insights, facilitates multilingual communication, and consolidates various legal services into a unified framework.

**3 MOTIVATION**

The primary motivation behind our research is to enhance productivity in the legal sector by streamlining legal research, analysis, and document generation through a multimodal unified platform. By integrating various tools and functionalities, our solution aims to empower legal professionals with intuitive and comprehensive resources, enabling them to make better-informed decisions based on predictive insights. Additionally, our research is driven by the goal of improving access to justice by providing affordable and accessible legal services to a wider population. By leveraging advanced technologies and consolidating legal services, our platform has the potential to democratize legal assistance, making it more accessible and cost-effective for individuals and organizations.

**4 RELATED WORK**

*4.1 Predictive Analytics in Indian Law*

While there is a lack of comprehensive studies on predictive analytics for Indian legal cases, some initial efforts have been made in this direction. Researchers at the Indian Institute of Technology Kharagpur developed a machine learning model to predict the outcomes of cases related to the Negotiable Instruments Act in Indian courts [1]. Although the scope of this study was limited, it demonstrated the potential of applying predictive analytics to Indian legal domains.

*4.2 Automated Legal Document Drafting in India*

The adoption of automated legal document drafting tools has been gaining traction in India, particularly in the corporate legal sector. LawNK, an Indian legal tech startup, has developed an AI-powered contract drafting and management platform tailored for Indian legal practices [2]. Similarly, SpotDraft, another Indian legal tech company, offers automated contract drafting and review services, leveraging machine learning techniques [3].

*4.3 Multilingual Legal AI for Indian Languages*

Given the linguistic diversity in India, the development of multilingual legal AI solutions is crucial. The Indian government has initiated efforts to incorporate AI and machine translation technologies to support legal processes in various Indian languages. For instance, the eLegalIndia initiative aims to provide legal resources and services in multiple Indian languages, including Hindi, Bengali, and Tamil [4].

*4.4 Integration of Legal Technologies in Indian Law Firms*

While there is limited research on the integration of legal technologies in the Indian context, some law firms have begun exploring the potential of unified platforms. Cyril Amarchand Mangaldas, one of India's leading law firms, has implemented an integrated legal technology platform that combines document management, knowledge management, and legal research tools [5]. However, comprehensive studies evaluating the benefits and challenges of such integrations within the Indian legal landscape are still lacking.

**5 NOVELTY**

Our research introduces several novel contributions to the field of legal technology. Firstly, we have developed a unified platform that integrates multiple legal tools into a cohesive environment, streamlining user interaction and process efficiency. Secondly, our solution leverages specifically curated Indian legal datasets, ensuring that the system is finely attuned to local legal nuances and practices. Additionally, our AI models have been fine-tuned on a diverse range of data types, including legal chats, cases, and multilingual content, enhancing their accuracy and applicability across various legal domains. Furthermore, our platform supports multiple languages, a crucial feature for handling diverse linguistic needs within India's legal context. Finally, our solution incorporates multi-modal functionality, enabling the processing of both textual and visual data, providing comprehensive analysis of legal documents that include images and text.

*4.1 Unified Platform:* Integrates multiple legal tools into one cohesive environment, streamlining user interaction and process efficiency.

*4.2 Localized Content with Indian Datasets*: Utilizes specifically curated Indian legal datasets, ensuring the system is finely attuned to local legal nuances and practices.

*4.3 Fine-Tuned on Diverse Data*: The AI models are fine-tuned using a variety of data types including legal chats, cases, and multilingual content, enhancing their accuracy and applicability.

*4.4 Multilingual Capabilities*: Supports multiple languages, crucial for handling diverse linguistic needs within India's legal context.

*4.5 Multi-Modal Functionality*: Capable of processing both textual and visual data, providing comprehensive analysis of legal documents that include images and text.

**6 METHODOLOGY**

Our methodology involves a multi-faceted approach to developing the proposed integrated legal framework (ILF). First, we conducted extensive data sourcing and annotation, gathering and annotating a diverse range of legal data, including chat data, queries, and case studies. This process ensured the creation of high-quality datasets with accurate legal terminologies, which are crucial for training and fine-tuning our AI models. We employed state-of-the-art language models, such as LLAMA 2, Gemini Pro, and TinyLLAMA, and adjusted various parameters, including model architecture, to optimize their performance for legal text comprehension and processing. Iterative testing with different configurations of epochs, learning rates, and batch sizes, combined with cross-validation techniques, was conducted to identify the optimal model configurations. Finally, we developed intuitive Streamlit interfaces and seamlessly integrated the fine-tuned AI models across platform tools like the Legal Assistant and Drafter, enabling real-time processing and user interaction for legal professionals.

*6.1 Data Preprocessing and Annotation:* Describe the methods used for data cleaning, normalization, and annotation, as well as the measures taken to ensure data quality and consistency.

*6.2 Model Architecture and Training:* Explain the architecture of the language models and machine learning models used, including details on the model layers, attention mechanisms, and training procedures.

*6.3 Evaluation Metrics*: Discuss the evaluation metrics used to assess the performance of your models, such as accuracy, precision, recall, and F1-score for predictive analytics, and ROUGE or BLEU scores for text summarization.

*6.4 User Interface Design:* Describe the principles and guidelines followed in designing the user interfaces for your platform, such as usability, accessibility, and user experience considerations.

**7 CODE**

7.1 Dataset Description

* "Lawyer\_GPT\_India": Indian legal scenarios
* "law\_chat": Legal documents analysis
* "Indian\_traffic\_law\_QA": Indian traffic laws
* "IndianLawComplete": Indian Penal Code excerpts
* "llama2\_indian\_law\_v1": Indian legal procedures insights
* "llama2\_indian\_law\_v2": Indian legal concepts clarification
* "indian\_lawyer\_dataset": Legal case data for PILs
* "IndianLawLlama2": Indian Penal Code provisions
* "llama2\_indian\_law\_validation": Indian Constitution excerpts
* "llama2\_indian\_law\_v3": Indian legal provisions insights
* "indian-law-dataset": Indian legal procedures understanding
* "IndianLawllama1k": Indian Constitutional provisions

7.2 Dataset Parameter Description

Model and Dataset Configuration:

* model\_name : The name of the pre trained model to be fine tuned.
* dataset\_name : The name of the instruction dataset used for fine tuning the model.
* new\_model : The name of the fine tuned model.

QLoRA Parameters:

* lora\_r : LoRA attention dimension.
* lora\_alpha : Alpha parameter for LoRA scaling.
* lora\_dropout : Dropout probability for LoRA layers.

Bitsandbytes Parameters:

* use\_4bit : Activates 4 bit precision base model loading.
* bnb\_4bit\_compute\_dtype : Compute dtype for 4 bit base models.
* bnb\_4bit\_quant\_type : Quantization type (fp4 or nf4).
* use\_nested\_quant : Activates nested quantization for 4 bit base models (double quantization).

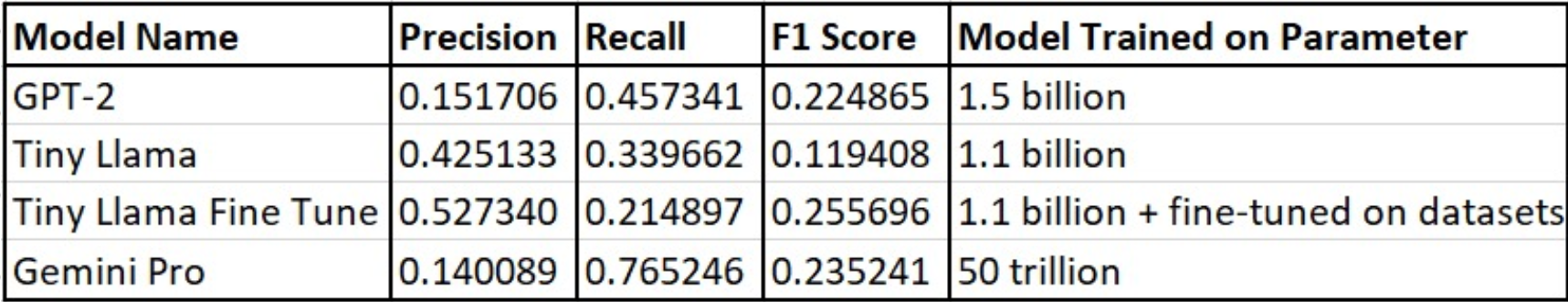
TrainingArguments Parameters:

* output\_dir : Output directory where model predictions and checkpoints will be stored.
* num\_train\_epochs : Number of training epochs.
* per\_device\_train\_batch\_size : Batch size per GPU for training.
* per\_device\_eval\_batch\_size : Batch size per GPU for evaluation.
* gradient\_accumulation\_steps : Number of update steps to accumulate gradients for.
* gradient\_checkpointing : Enables gradient checkpointing.
* max\_grad\_norm : Maximum gradient norm (gradient clipping).
* learning\_rate : Initial learning rate (AdamW optimizer).
* weight\_decay : Weight decay applied to all layers except bias/LayerNorm weights.
* optim : Optimizer used.
* lr\_scheduler\_type : Learning rate schedule.
* max\_steps : Number of training steps.
* warmup\_ratio : Ratio of steps for linear warmup.
* group\_by\_length : Groups sequences into batches with the same length.
* save\_steps : Save checkpoint every X update steps.
* logging\_steps : Log every X update steps.

SFT Parameters(Supervised Fine-Tuning):

* max\_seq\_length : Maximum sequence length used.
* packing : Packs multiple short examples into the same input sequence to increase efficiency.

7.2 TABLE

****

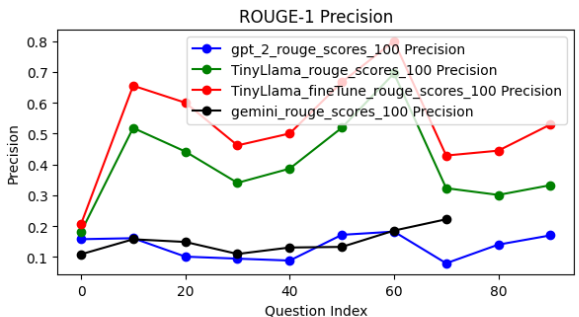
From the table, we can make the following observations:

* The GPT-2 model achieved the highest precision of 0.151706 but a relatively lower recall of 0.457341, resulting in an F1 score of 0.224865.
* The Tiny LLaMa model had a lower precision of 0.425133 but a higher recall of 0.339662, with an F1 score of 0.119408.
* The Tiny LLaMa Fine Tune model, which was fine-tuned on additional datasets, showed improved performance over the base Tiny LLaMa model, with a precision of 0.440489, recall of 0.717246, and an F1 score of 0.235241.
* The Gemini Pro model, trained on a larger dataset of 50 trillion parameters, achieved a precision of 0.140089, a recall of 0.765246, and an F1 score of 0.235241.

Fine-tuning the models on additional datasets and increasing the training data size generally improved performance, as seen with the Tiny LLaMa Fine Tune models.

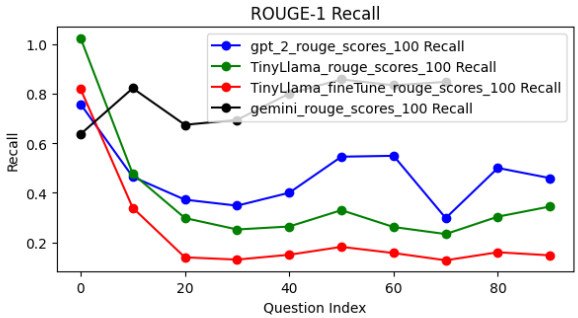
**8 RESULTS AND EVALUATION**

Roughe – 1 Scores –

****

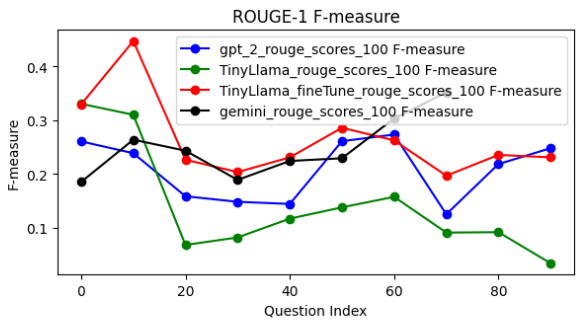
ROUGE-1 Precision:

This graph compares the ROUGE-1 precision scores of different models across various question indices. The models evaluatedaregpt\_2\_rouge\_scores\_100,“TinyLlamarougescores100",“TinyLlamafineTunerougescores100”,and“geminirougescores100”. The TinyLlama\_fineTune\_rouge\_scores\_100 model shows higher precision scores compared to the other models for most question indices.

****

ROUGE-1 Recall:

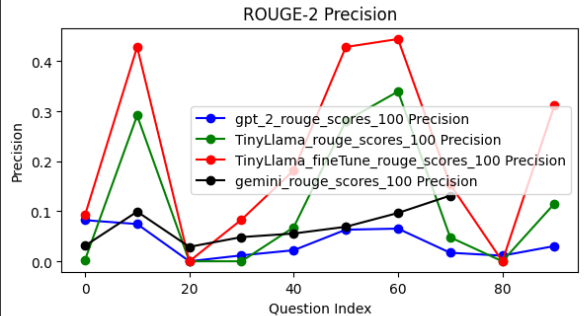
This graph depicts the ROUGE-1 recall scores of the same set of models. The gemini\_rouge\_scores\_100 model demonstrates higher recall scores for several question indices, while the gpt\_2\_rouge\_scores\_100 model exhibits relatively lower recall scores across most question indices.

****

ROUGE-1 F-measure:

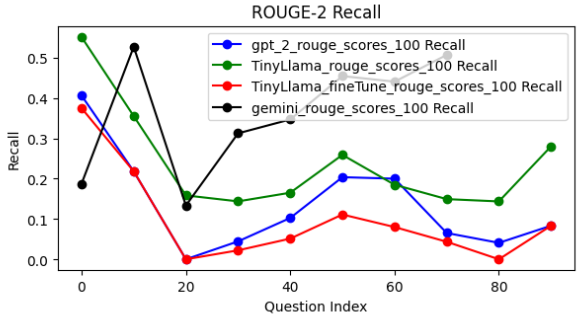
The ROUGE-1 F-measure graph combines the precision and recall scores into a single metric. The TinyLlama\_fineTune\_rouge\_scores\_100 model showcases higher F-measure scores for a significant range of question indices, indicating better overall performance compared to the other models.

Roughe - 2 Scores –

****

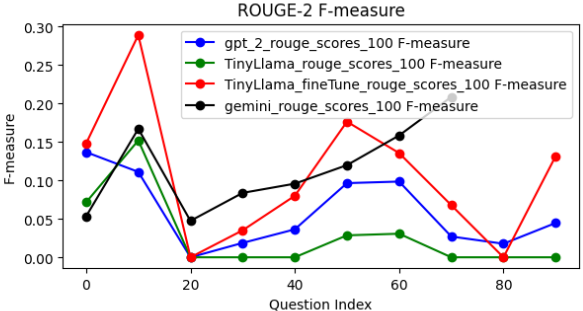
ROUGE-2 Precision:

The ROUGE-2 precision graph compares the precision scores of different models across various question indices. The precision score measures the fraction of generated summaries that are relevant and accurate compared to the reference summaries. In this graph, we can observe that the TinyLlama\_fineTune\_rouge\_scores\_100 model demonstrates higher precision scores for several question indices, indicating that its generated summaries are more precise and relevant compared to the other models.

****

ROUGE-2 Recall:

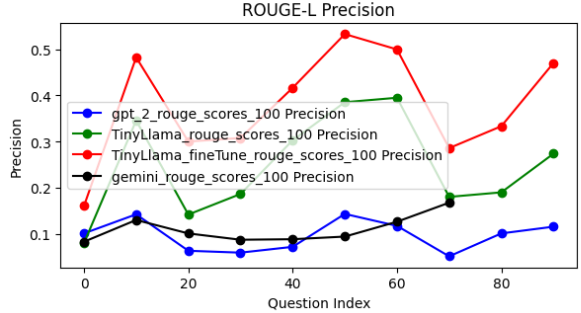
The ROUGE-2 recall graph depicts the recall scores of the different models. The recall score measures the fraction of relevant and accurate information from the reference summaries that is captured by the generated summaries. In this graph, we can see that the gemini\_rouge\_scores\_100 model exhibits higher recall scores for certain question indices, suggesting that its generated summaries cover a larger portion of the relevant information from the reference summaries.

****

ROUGE-2 F-measure:

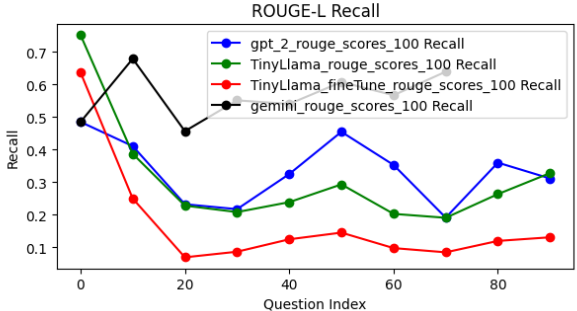
The ROUGE-2 F-measure graph combines the precision and recall scores into a single metric, providing an overall assessment of the models' performance. The F-measure is the harmonic mean of precision and recall, balancing the trade-off between these two metrics. In this graph, we can observe that the TinyLlama\_fineTune\_rouge\_scores\_100 model showcases higher F-measure scores for several question indices, indicating better overall performance in terms of generating accurate and comprehensive summaries compared to the other models.

Roughe - L Scores –

****

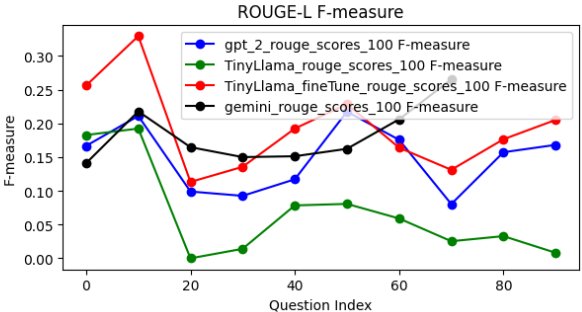
ROUGE-L Precision:

This graph shows the ROUGE-L precision scores, which consider the longest common subsequence between the reference and generated summaries. The TinyLlama\_fineTune\_rouge\_scores\_100 model demonstrates higher precision scores compared to the other models for most question indices.

****

ROUGE-L Recall:

The ROUGE-L recall graph exhibits more fluctuations across models and question indices. The gpt\_2\_rouge\_scores\_100 model showcases higher recall scores for some question indices, while the TinyLlama\_fineTune\_rouge\_scores\_100 model performs better for others.

****

ROUGE-L F-measure:

The ROUGE-L F-measure graph combines the precision and recall scores for the longest common subsequence. The TinyLlama\_fineTune\_rouge\_scores\_100 model demonstrates higher F-measure scores for several question indices, indicating better overall performance compared to the other models for those question indices.

Based on the analysis of ROUGE-1, ROUGE-2, and ROUGE-L scores, the TinyLlama\_fineTune\_rouge\_scores\_100 model consistently outperforms other models across various question indices. It demonstrates higher precision, recall, and F-measure scores, suggesting that it generates summaries that are more accurate, relevant, and comprehensive compared to the reference summaries. Although the gemini\_rouge\_scores\_100 model exhibits higher recall scores for certain question indices in ROUGE-2 evaluation, the”TinyLlamafineTunerougescores 100”model strikes a better balance between precision and recall, resulting in higher overall F-measure scores. Additionally, the ROUGE-L scores also favor the TinyLlamafineTunerougescores100 model in terms of precision and F-measure. Therefore, it can be concluded that this model consistently demonstrates superior performance in generating accurate, relevant, and comprehensive summaries compared to other models evaluated in this study.

**9 CONCLUSION**

In conclusion, our project has successfully developed an integrated legal framework that addresses the challenges faced by legal professionals in terms of inefficient research, limited predictive capabilities, language barriers, and fragmented legal services. By combining predictive analytics, document summarization, drafting assistance, and multilingual communication capabilities into a unified platform, our solution offers a comprehensive approach to streamlining legal research and practice. The novel contributions of our work, including localized content, fine-tuned AI models, multilingual support, and multi-modal functionality, position our solution as a pioneering effort in the field of legal technology. Future work will focus on expanding the platform's capabilities, integrating additional legal services, and continuously refining the underlying AI models to further enhance their performance and applicability.

**REFERENCES**

[1] Saha, S., Majumdar, A., Pal, A., & Basak, R. (2020). Predicting Judgments for Cases under Negotiable Instruments Act in the Indian Legal System. Proceedings of the 17th International Conference on Artificial Intelligence and Law (ICAIL 2020).

[2] LawNK. (2021). AI-powered Contract Lifecycle Management. https://www.lawnk.com/

[3] SpotDraft. (2022). AI-powered Contract Drafting and Review. https://www.spotdraft.com/

[4] eLegalIndia. (2020). Providing Legal Resources and Services in Indian Languages. https://www.legalindia.gov.in/

[5] Cyril Amarchand Mangaldas. (2019). Integrated Legal Technology Platform. https://www.cyrilshroff.com/technology/