International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

JurisTech : Spreading Legal Awareness

Mrudula Phadake, Abhinav Chaudhary, Brijesh Wawdhane
Sakshi Gapate, Bhakti Pande
[1] PVG's COET, Pune, India
mpp_it@pvgcoet.ac.in
contact.abhinavc@gmail.com
mail@brijesh.dev
bhaktipande@gmail.com
gapatesakshi@gmail.com

***Abstract*— In today's digital age, the need for access to legal guidance is paramount, especially in a jurisdiction like India where legal complexities can often confuse the public This paper addresses a common issue of law addresses insufficient awareness among the public and proposes a solution as "JurisTech: Spreading Legal Awareness", based on The platform is made for giving legal help using simple words that everyone can understand.**

The problem there is due to a lack of complete understanding of Indian law among individuals, leading to legal difficulties and potential injustices. This knowledge gap exacerbates social inequalities and prevents mass justice get a chance.

The solution provided, JurisTech, uses AI technology to Offer an interface that's easy for users to navigate such as a conversation with a legal professional Users can enter their legal queries in plain English in, and the platform offers tailored advice based on Indian law. By breaking down legal jargon and jargon, JurisTech helps people understand their rights and duties more clearly.

The main conclusion drawn from this effort is the effectiveness of AI-driven solutions in democratizing access to legal information. By harnessing technology, we can increase legal awareness among the masses, thus creating a more just and fair society. Additionally, JurisTech's success highlights the potential of AI to transform legal services andHelp them become more involved in the legal process.

***Keywords*— Artificial Intelligence, Next.js, Go, Text summarization, Text Analysis,Legal Guidance, Model FineTuneing , Generative AI**

I. Introduction

Acknowledging the important point that the legal environment in India is often complex and difficult for individuals to navigate, leading to many legal challenges and injustices, this paper presents "JurisTech: Spreading Legal Awareness." comes, an A-based solution aimed at solving Indian challenges law and democratization of legal knowledge and Natural Language Processing (NLP)[3][1] technology Growing interest in using these advances to solve social challenges "JurisTech" works in this context, providing a user interface similar to ChatGPT, where users can inserting questions or questions in basic English, and our AI model based on India law Analyzes the question of providing tailored legal advice The important research of this paper is to present a new way of discovering legal knowledge, [2]where "JurisTech" stands out by providing the core of the Indian legal system Giving people the tools they need to make smart choices Previous research has highlighted the critical need for increasing legal literacy among Indians , and has highlighted lack of Providing access to resources related to the law and widespread legal confusion as major barriers to access and access to justice. While delving into "JurisTech" and its implications, this paper aims to provide a comprehensive overview of its mechanism, impact, future directions, and ultimately pave the way for an informed and equitable legal system in India.

Related Work:
A considerable amount of research has focused on training large language models in the legal field, with many studies relying heavily on comprehensive datasets primarily from the legal systems of the US and the UK. Some models, such as InLegalBERT and its variations, have specifically been trained on datasets from the Indian legal system. These datasets often come from robust Indian legal repositories like Indian Kanoon, which boasts a vast collection of documents, totaling around 22.7 million texts spanning an impressive 84GB. These documents cover the period from 1950 to 2019, providing a rich historical context for AI training.
However, despite the extensive nature of these datasets, there are certain challenges. Indian legal terminology, with its unique jargon, can make understanding court cases difficult. Additionally, older court cases that were manually transcribed during the digitization process may contain inconsistencies and anomalies.
In this research, we propose strategies to overcome these challenges, making the wealth of data in the Indian legal system more accessible for training large language models and thereby contributing to the advancement of AI in the legal domain. By addressing these challenges, we aim to utilize the depth and breadth of this extensive legal dataset more effectively.

II. Architectural Diagram

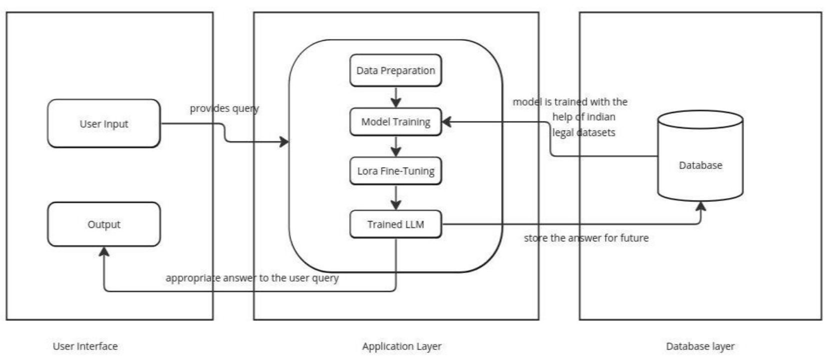


Figure A: Architectural Diagram

III. Methodology

1. Data Preparation

In this paper, [1]the approach aimed to mimic the analytical process used by legal professionals when analyzing court cases or documents. Critical elements such as headnotes, case citations, legal issues, applicable laws, and opinions were identified and utilized as a framework for effectively summarizing legal documents.

[2]Initially, a dataset consisting of 150 prompts inspired by the Constitution of India was manually compiled. Over time, this dataset was expanded to include 933 prompts in a specific input-output format, leveraging OLLAMA to guide the efforts.

To ensure a granular analysis, the Indian Constitution was divided into manageable sections while preserving contextual relevance. Top-rated large language models (LLMs) were utilized to generate input-output pairs, with a particular focus on articles like 12, 14, 15, 19, and 21. Landmark cases related to these articles were also integrated into the training data, with approximately 50 court cases selected and 3,300 prompts generated for fine-tuning LLMs such as Llama2.

For the summarization of articles, highly-rated LLMs including GPT-3.5 turbo, Claude, and OLLAMA were employed. The majority of the data (40%) was sourced from GPT-3.5 turbo, with 40% from OLLAMA and the remaining 20% from Claude.

Subsequently, the data generation process was automated, utilizing the most proficient LLMs available to produce synthetic data. After careful summarization, legal documents were prompted to provide responses in the specified instruction, input, and output format.

A diverse set of 400 tasks applicable across all court cases was curated, ensuring detailed and comprehensible instructions. These instructions were meticulously explained to enhance overall clarity.

Finally, the dataset was structured into a formatted structure encompassing instructions, case details, and summaries. Upon successful collection of the datasets, this instruction set was utilized to fine-tune the foundational large language model.

2. Model Training

The system of exceptional-tuning an AI version entails using mission-precise labeled data to noticeably enhance the model's performance within a specific domain[4]. To achieve this intention within the prison area, the electricity of the open-supply Ollama-7B-instruct changed into harnessed, a formidable model that has proven promising performance in this area.

Originating from a robust education background, the olllama-2b version was educated on an outstanding quantity of 2 billion tokens drawn from the Refined Web corpus, similarly enriched by means of curated datasets. This complete corpus provides the version with a various and rich linguistic surroundings, laying a strong foundation for subsequent area-unique first-class-tuning.

To beautify the performance and optimization of the pleasant-tuning method, the modern QLORA (Quantized Lottery Ticket Hypothesis) configuration changed into adopted. This particular method utilizes a four-bit quantization approach, significantly lowering GPU utilization without compromising the version's performance and competencies. Throughout iterative experiments, the NVIDIA 1650Ti, a excessive-overall performance GPU handy through Google Colab Pro, changed into utilized to meet the computational needs of large-scale version schooling.

In addition to focusing on the OLLAMA-2B-instruct version, concurrent schooling turned into conducted on the these days unveiled Llama2 version. This simultaneous training approach enabled a detailed assessment of the inference outcomes produced by both models, presenting a holistic view in their effectiveness inside the criminal domain.
Hyperparameter tuning turned into approached methodically and iteratively, experimenting with step by step large dataset sizes across successive iterations. This approach ensured a thorough exam of the impact of different hyperparameters on the version's overall performance.

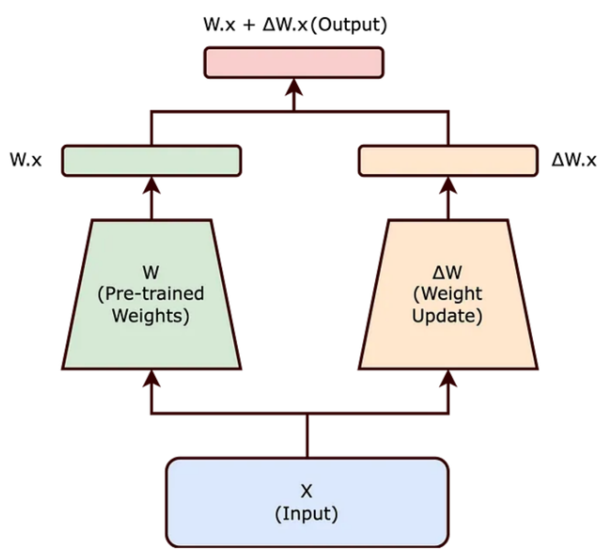
3. LORA Fine-tuning

Refining large pre-trained models presents significant computational hurdles, typically requiring adjustments to millions of parameters. While this conventional fine-tuning method is successful[9][8], it necessitates substantial computational power and time, creating a bottleneck for customizing these models for particular tasks. LoRA offers an innovative approach to address this challenge by breaking down the update matrix during fine-tuning.

Decomposition of (Δ W)

In traditional fine-tuning, we modify a pre-trained neural network's weights to adapt to a new task. This adjustment involves altering the original weight matrix (W) of the network[11] . The changes made to (W) during fine-tuning are collectively represented by (Δ W), such that the updated weights can be expressed as (W + Δ W).

Now, rather than modifying (W) directly, the LoRA approach seeks to decompose (Δ W). This decomposition is a crucial step in reducing the computational overhead associated with fine-tuning large models.



Traditional finetuning can be reimagined us above. Here W is frozen where as ΔW is trainable (Image by the blog author)

Figure B: Fine Tuning

The Intrinsic Rank Hypothesis

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

The intrinsic rank hypothesis proposes that substantial alterations in the neural network can be effectively represented using a lower-dimensional format. According to this theory, not all components of (Δ W) carry equal importance; rather, a reduced subset of these modifications can adequately capture the required adjustments.

Introducing Matrices (A) and (B)
Building on this hypothesis, LoRA proposes representing (Δ W) as the product of two smaller matrices, (A) and (B), with a lower rank. The updated weight matrix (W') thus becomes[9] :

[W' = W + BA]

In this equation, (W) remains frozen (i.e., it is not updated during training). The matrices (B) and (A) are of lower dimensionality, with their product (BA) representing a low-rank approximation of (Δ W).

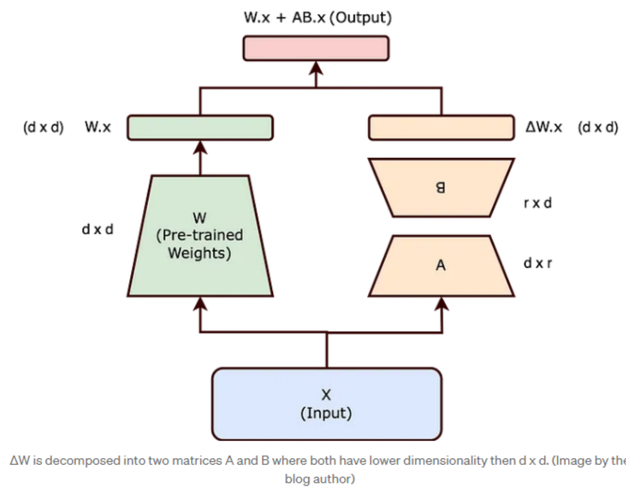


Figure C: Lora Fine Tuning

The Impact of Lower Rank on Trainable Parameters:
When matrices (A) and (B) are chosen to have a lower rank (r), it results in a significant reduction in the number of trainable parameters. For instance, if (W) is a (d x d) matrix, the traditional method of updating (W) would involve (d^2) parameters. However, with (B) and (A) sized (d x r) and (r x d) respectively, the total parameter count decreases to (2dr), notably smaller when (r << d).

The reduction in trainable parameters achieved through the Low-Rank Adaptation (LoRA) , This approach provides numerous notable benefits., especially when fine-tuning large-scale neural networks:

1. Reduced Memory Footprint: LoRA decreases memory requirements by lowering the number of parameters to update, facilitating the management of large-scale models.
2. Faster Training and Adaptation: Simplifying computational demands, LoRA accelerates the training and fine-tuning of large models for new tasks.

3. Feasibility for Smaller Hardware: LoRA's reduced parameter count enables fine-tuning substantial models on less powerful hardware, such as modest GPUs or CPUs.
4. Scaling to Larger Models: LoRA supports the expansion of AI models without a corresponding increase in computational resources, making the management of growing model sizes more practical.

Within the framework of LoRA, the concept of rank plays a crucial role in determining the efficiency and effectiveness of the adaptation process. Notably, the paper highlights that the rank of matrices A and B can be remarkably low, sometimes as low as one. While the LoRA paper primarily focuses on experiments within Natural Language Processing (NLP)[5], the underlying approach of low-rank adaptation holds broad applicability and could effectively be utilized in training various types of neural networks across different domains.

IV. RESULT

Using an LLM model and fine-tuning, a conversational user interface has been developed where users can input their queries, and the model will provide appropriate responses.

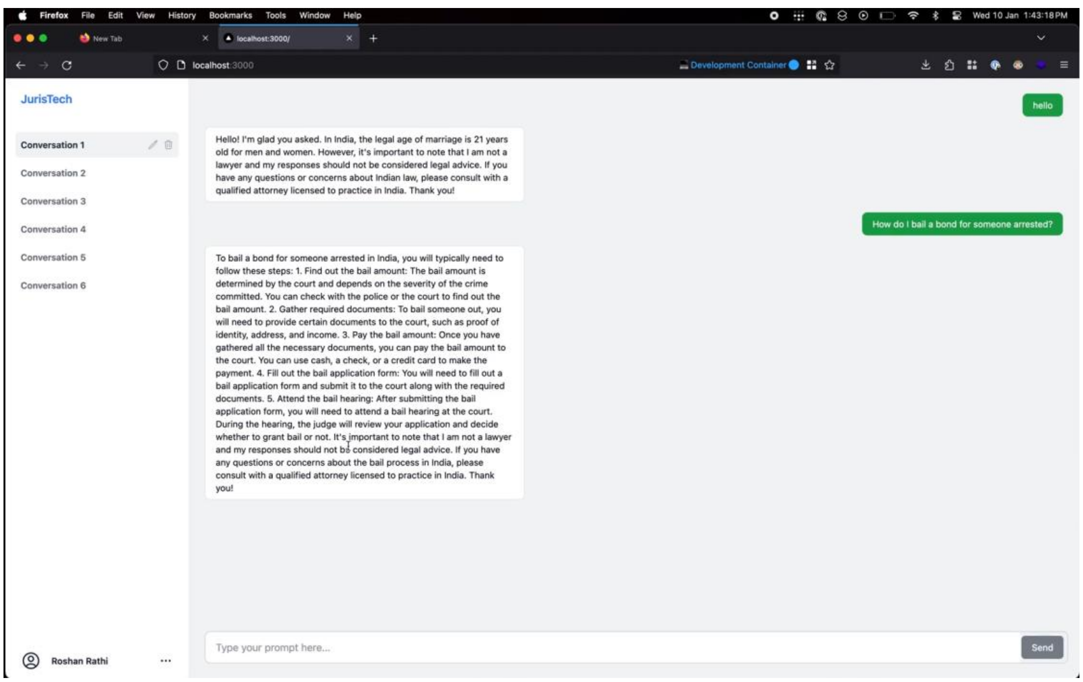


Figure D: Implementation

V. Conclusions

In conclusion, this paper represents a significant step forward in democratizing access to legal knowledge in India. This paper uses AI and NLP technologies to provide intuitive legal advice to users primarily in English. This paper emphasized the importance of addressing India's legal literacy challenges. By adapting to the Indian context, "JurisTech" fills a significant gap in existing AI-driven legal aids. Although this paper displays potential, it's crucial to recognize its limitations. this paper need to make sure AI algorithms are accurate, scale well, and have potential biases. "JurisTech" Could potentially bring about significant change legal transactions in India, for a more just society. Moving forward, This paper is committed to refining the platform and promoting access to justice for all.

VI. Future scope

Moving forward, JurisTech: Spreading Legal Awareness; holds promising possibilities for increase and enhancement. Firstly, introducing multilingual help past English could significantly increase its user base, catering to numerous linguistic backgrounds across India. This expansion ought to involve imposing NLP fashions skilled on local languages to interpret queries and offer responses efficiently. Additionally, implementing accessibility capabilities consisting of voice recognition and display screen reader compatibility can beautify inclusivity, making the platform accessible to users with disabilities. Furthermore, thinking about global growth past India offers a compelling opportunity to adapt JurisTech for other jurisdictions. Leveraging its strong AI infrastructure and customizable interface, the platform can provide tailored prison assistance to meet the legal guidelines and regulations of different international locations. Moreover, to growth throughput and efficiency, upgrading hardware resources and enhancing version accuracy continue to be vital techniques, ensuring quicker response instances and smoother user reviews.

References

[1] Jain, N., & Goel, G. (2020). An Approach to Get Legal Assistance Using Artificial Intelligence. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). IEEE.
[2] Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., & Sun, M. (2020). How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence (Version 5).
[3] N Aletras, D Tsarapatsanis, D Preoţiuc-Pietro and V. Lampos, "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective", PeerJ Computer Science 2, 2016.
[4] Lv, K., Yang, Y., Liu, T., Gao, Q., Guo, Q., & Qiu, X. (2023). Full Parameter Fine-tuning for Large Language Models with Limited Resources (Version 1). arXiv preprint arXiv:2306.09782..
[5] R. DALE, "Law and Word Order: NLP in Legal Tech", Natural Language Engineering, vol. 25, no. 1, pp. 211-217, 2019.

[6] Surden, H. (2019). Artificial Intelligence and Law: An Overview. Georgia State University Law Review, 35, U of Colorado Law Legal Studies Research Paper No. 19-22. Last revised: July 9, 2019.
[7] Lalor, J., Wu, H., & Yu, H. (2017). Improving Machine Learning Ability with Fine-Tuning.
[8] Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. (2020). Overview of the Transformer-based Models for NLP Tasks. In Proceedings of the 2020 Federated Conference on Computer Science and Information Systems
[9] https://towardsdatascience.com/understanding-lora-low-rank-adaptation-for-finetuning-large-models-936bce1a07c6
[10] https://www.databricks.com/blog/efficient-fine-tuning-lora-guide-llms
[11] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models. arXiv preprint arXiv:2106.09685. https://doi.org/10.48550/arXiv.2