



Indian Institute of Technology Delhi

School of Interdisciplinary Research (SIRe)

Project Proposal for Ph.D.



SIRe Project Number

2024-01-39

Project Details

Project Title	Legal Applications of AI (LLMS)
---------------	---------------------------------

Project Summary	The legal domain is known to be complex, slow, expensive and only accessible to the privileged few. This is particularly true in Indian context where vast segment of the population is often deprived of the legal facilities due to lack of awareness, money, education, etc. The aim of the project is to extend accessibility to legal processes to the vast majority through development of AI-based chatbot that will be helpful to the masses. It should be able to assist access to justice by speeding up legal processes and reducing legal costs. These obstacles exist due to inherent complexity in the legal domain which current language models struggle to capture. Our aim is to develop novel natural language processing and machine learning algorithms which can assist in legal advice, legal research, document analysis, and mining insights. This project will focus on the following areas: • Development of a Legal Language Model (LLM). Legal texts are complex and hard to understand for standard language models. Legal questions require complex reasoning and the ability to process if-then scenarios. To tackle this issue, we plan to create a specialized Legal Language Model (LLM) specifically designed for the legal system. This model will be trained to understand the unique structure and content of legal documents, such as court opinions and statutes. By teaching the LLM to recognize different types of legal information and relationships within them, we aim to improve the accuracy and usefulness of legal AI tools for researchers and practitioners. • Development of data anonymization techniques. The aim is to anonymize personally Identifiable Information (PII), to maintain privacy. For example, chat and call center transcripts which require analysis for informing government policies may contain names, contact details, and descriptions of their legal issues, which need to be anonymized in such a way that the informative contents may be used for the purpose of designing the ML system without revealing any individual information. Appropriate schemes for anonymization have to be developed so that the extent of anonymization can be decided automatically depending upon the requirement. • Development on an LLM that asks questions. The aim is to optimize pre-consultation efficiency, enabling the chatbot to gather pertinent data from citizens prior to legal consultation or litigation. For example, if a user seeks advice on a potential personal injury claim, the chatbot should be able to ask about the relevant background, such as circumstances of the incident, any medical treatment received, among others. • Development of a Q&A system through controlling hallucination. AI-generated legal information may sometimes provide inaccurate or misleading responses, leading to potential legal risks and misinformation. To address this challenge, we aim to develop a Question & Answer (Q&A) system that effectively controls hallucination. By integrating knowledge bases, implementing answer justification mechanisms, and conducting factuality checks, our system will confine AI responses within the bounds of verified legal knowledge. This approach aims to minimize the risk of providing inaccurate information, upholding the highest standards of accuracy and reliability in legal AI applications. Application of AI and ML techniques to the legal domain has the potential to create a system that can speed up justice, reduce backlogs, reduce the spread of misinformation, increase the trust of citizens in law and order. Hence, the overall goal of the project is to provide equal access to justice for all in India, without compromising users' privacy.
------------------------	---

PhD Supervisors			
Role	Faculty	Academic Unit in IITD	Email ID
Supervisor 1	Prof. Jyoti Kumar Associate Professor	Department of Design	jyoti@design.iitd.ac.in
Supervisor 2	Prof. Aparna Mehra	Department of Mathematics	apmehra@maths.iitd.ac.in

Project requirements (Student qualifications, experience required, etc.)
• B.Tech in Computer Engineering/ Computer Sc and related areas (with at least 65%) • Masters in Comp Sc. with specialization if AI-ML/Pattern Recognition. • At least two years experience in the fields of search engines, analytics, ML. • Familiarity with legal domain will be preferable.

Source of funding (IRD/FITT Project details, if any)
Preferably candidate with own funding on part-time basis.

Role of Faculty Members involved
Role of PI: Guiding in training a legal LLM, AI-ML based modelling the data anonymization problem and development of AI algorithms, Developing AI-ML techniques to reduce hallucinations. Role of Co-PI: Analysis of the needs of the three identified categories of the users' (lawyers, advise seekers and admin) in the legal domain, Design of interactive systems that combine strengths of

human and machine during data anonymization, hallucinations, case review, and document generation.

