

**MADRAS INSTITUTE OF TECHNOLOGY
ANNA UNIVERSITY
CHENNAI – 600 044.**



**Project Proposal
on
“AI-Driven Legal Document Research Assistant
with
Enhanced Context Understanding”**

**Submitted by
Dr. M. R. Sumalatha,
Professor**

**Department of Information Technology,
Madras Institute of Technology Campus,
Anna University, Chennai.**

PROJECT PROPOSAL

Project Title:

AI-Driven Legal Document Research Assistant with Enhanced Context Understanding

Project Overview:

The AI-Driven Legal Document Research Assistant with Augmented Context Understanding aims at addressing legal information overflow through the deployment of intelligent, context-sensitive judges/lawyers' assistant that facilitates fastening/drafting research without dislodging judicial discretion. Keyword search conventionally misses out on subtle legal semantics, yet this system exploits retrieval-augmented generation, multi-level attention, and graph-based reasoning for effective inter-connections between facts, statute, and precedent. Automated case summarization, retrieval of precedents/statutes, generation of draft orders, and explainable reasoning are presented within a judge-in-the-loop paradigm for the sake of transparency, retainability, and control. Modular toolkits with specificities for certain domains such as criminal law, cybercrime, financial offenses, constitutional, environmental, are deployable for adjusting the assistant according to divergent judicial requirements. Decreasing time for research, facilitating consistency for decision-takings, enabling structured, citation-ready output, the overall goal for the project remains judicial efficiency enhancement, knowledge relevance facilitation, case backlog reduction for the judicial system.

Core Technologies to be used:

AI & NLP Models: InLegalBERT, Longformer, LLaMA-2/3 fine-tuned on Indian judgments, Sentence-BERT for embeddings.

Retrieval Systems: FAISS, Milvus, or Weaviate for dense vector search; BM25/Elasticsearch for lexical retrieval.

Knowledge Graph Tools: Neo4j, NetworkX, DGL (Deep Graph Library), PyTorch Geometric for heterogeneous graph reasoning.

Attention Mechanisms: Transformers (Hugging Face), Multi-Head Attention, Cross-Attention, Graph Attention Networks (GAT).

Document Handling: PyMuPDF, pdfplumber, Apache Tika, Tesseract OCR, spaCy for entity extraction (acts, sections, parties).

Backend Frameworks: FastAPI, Flask (Python) for ML services; Node.js + Express + Drizzle ORM for orchestration; gRPC for microservices.

Frontend / Judge Console: React + TypeScript with TailwindCSS or Material UI; Streamlit/Gradio prototypes for rapid testing.

Databases: PostgreSQL/MySQL for structured metadata; MongoDB for unstructured case notes; Redis for caching queries.

Search & Indexing: Elasticsearch, OpenSearch, or Meilisearch for statute/judgment

retrieval pipelines.

DevOps & Deployment: Docker, Kubernetes, Helm charts, CI/CD (GitHub Actions/GitLab CI), Nginx/Traefik for routing.

MLOps & Observability: MLflow, DVC (Data Version Control), Prometheus + Grafana for monitoring, Sentry for error tracking.

Cloud/Infra: On-prem GPU clusters (NVIDIA A100/V100), or hybrid cloud (AWS S3, EC2, Lambda, SageMaker / Azure ML / GCP AI Hub).

Security: TLS/SSL encryption, OAuth2/SSO for authentication, RBAC for judge/lawyer access, HashiCorp Vault for secrets.

Explainability & Auditing: SHAP, LIME, Captum for model explainability; audit logs stored in blockchain-inspired append-only ledgers for trust.

Collaboration Tools: WebRTC or socket-based live transcript streaming; integration with e-Courts APIs.

The project combines the following major components to produce a robust, judge-in-the-loop legal research assistant. At its core, it uses state-of-the-art AI and NLP architectures like InLegalBERT, Longformer, and fine-tuned transformers to distill the subtle semantics of legal language, so that the system is not just capturing surface-level keyword-based contexts but deeper ones as well. These are complemented with retrieval modules such as FAISS, Milvus, and Elasticsearch so that the relevant statutes and precedents are quickly returned accurately. Accompanying this is a knowledge graph infrastructure constructed using such tools as Neo4j and PyTorch Geometric, which inscribes the fine-grained relationships between facts, laws, parties, and evidence so that judges and lawyers have a better time dealing with multi-layered cases. To fortify reasoning, the project resorts to multi-level attention mechanisms like cross-attention and graph attention networks which correspond case facts with statutory provisions and precedents while retaining contextual deeper-ness.

On the implementation side, the project combines document processing pipelines with PyMuPDF, pdfplumber, and Tesseract OCR for deriving structured information out of case files, judgments, and evidences. The backend features FastAPI, Flask, for machine learning services, backed by Node.js, Express, for orchestration, while databases such as PostgreSQL, MongoDB, hold structured and unstructured legal information. For the user-facing judge console, technologies such as React, TypeScript, Streamlit create an interactive environment with modules for summaries, precedent investigation, draft orders, and explainability panels. The system is deployed securely with Docker, Kubernetes, MLflow, with observability covered under Prometheus, Grafana, compliance covered through role-based access, audit trails, encryption. Explainability is driven through SHAP, Captum, retrieval-linked rationales, such that the assistant functions transparently as a support for judges rather than supplanting judicial discretion.

Goals & Objectives:

Goals

- Develop a **judge-in-the-loop AI assistant** that accelerates legal research and drafting without replacing judicial discretion.
- Reduce the **time and effort** required by judges and lawyers to analyze lengthy case documents, statutes, and precedents.
- Enhance **consistency and fairness** in judicial decision-making by highlighting binding precedents and advisory sentencing ranges.
- Ensure **transparency and explainability** in AI outputs so that recommendations are traceable to statutes and judgments.
- Build **modular support** for different domains of law (criminal, cybercrime, financial, civil, constitutional, environmental, etc.).
- Provide a **secure, scalable, and auditable platform** that integrates smoothly with existing court systems.
- Contribute to reducing the **judicial backlog** by improving efficiency in case handling and drafting.
- Enable **real-time hearing support** by integrating speech-to-text and instant retrieval of relevant statutes and precedents during court proceedings.
- Incorporate **bias detection and fairness monitoring** to ensure equitable outcomes across different case types and social contexts.
- Provide **data-driven insights and analytics** on case trends, conviction rates, and sentencing patterns to assist judges, researchers, and policymakers in informed decision-making.

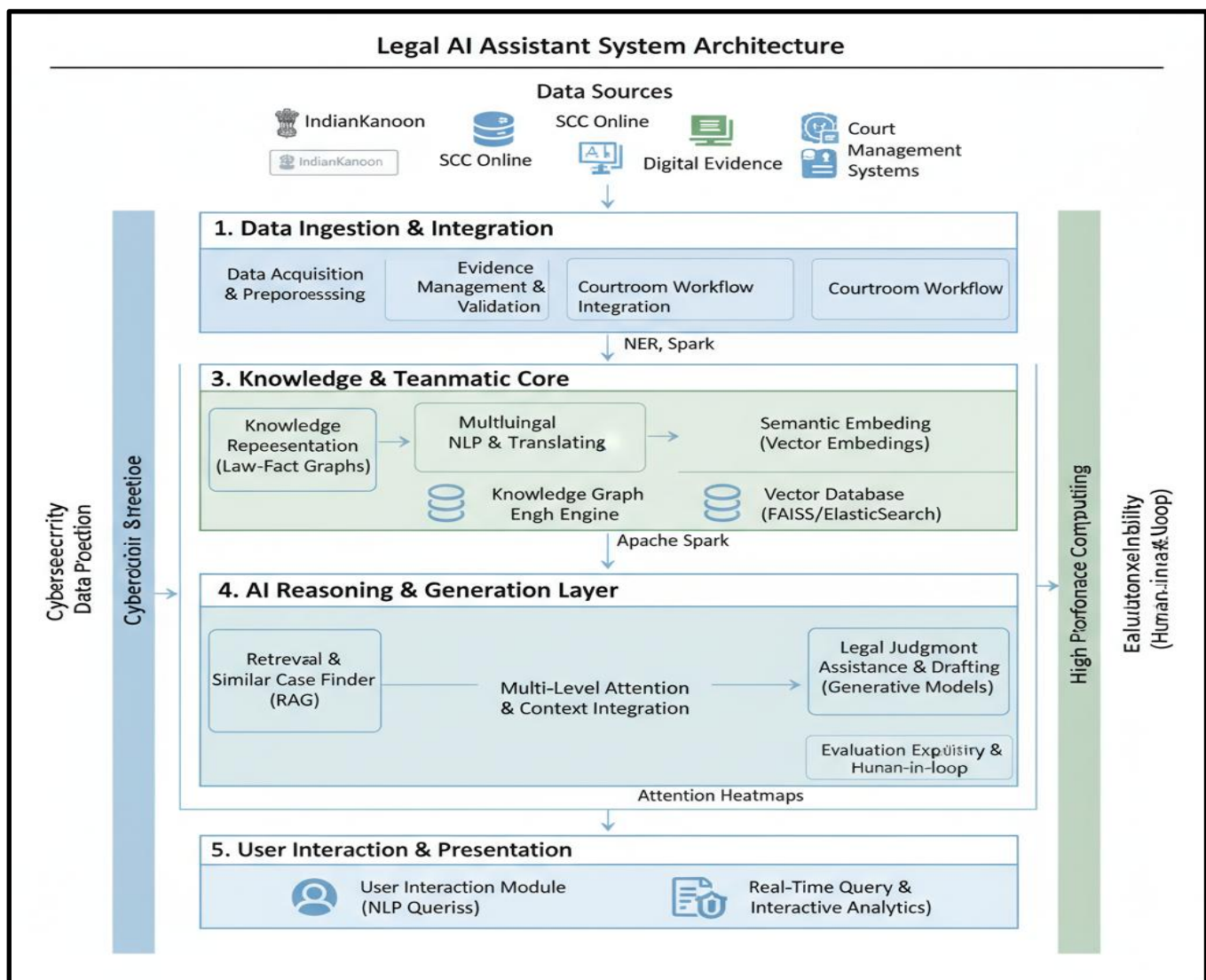
Objectives

- **AI & NLP Models:** Fine-tune domain-specific models like InLegalBERT, Longformer, and transformers to capture nuanced legal semantics.
- **Retrieval Systems:** Implement Retrieval-Augmented Generation (RAG) with FAISS/Elasticsearch to fetch statutes and precedents quickly.
- **Knowledge Graphs:** Construct heterogeneous law–fact graphs linking parties, statutes, evidence, and judgments for contextual reasoning.
- **Attention Mechanisms:** Apply multi-level attention, cross-attention, and graph attention networks to align facts with laws and precedents.
- **Document Processing:** Build pipelines using PyMuPDF, pdfplumber, and OCR for extracting structured information from case files.
- **Judge Console:** Design a user-facing interface (React/TypeScript) with modules for case summaries, precedent exploration, draft generation, and explainability.
- **Draft Order Generator:** Automate structured draft orders (bail, sentencing, civil

decrees) with editable outputs for judges.

- **Consistency Checker:** Develop tools to flag conflicts with binding precedents or outdated statutes.
- **Case-Type Modules:** Create specialized modules for domains like NDPS (Narcotics), Cybercrime, Financial Crimes, Terrorism, Family Law, and Constitutional cases.
- **Explainability Engine:** Integrate SHAP, Captum, and retrieval-linked rationales to ensure transparent, judge-controllable outputs.
- **Deployment & Security:** Use Docker, Kubernetes, MLflow, and RBAC to ensure secure, scalable deployment with audit logs.
- **Evaluation Metrics:** Measure efficiency (time saved), precision@k in retrieval, judge edit distance in drafts, and citation correctness.
- **Governance:** Maintain model registries, dataset versioning, and audit trails to ensure accountability and trust.

Architecture Diagram - Legal Research Assistant



MODULE WISE DESCRIPTION:

1. Data Acquisition & Preprocessing

This module constructs the ground for the system by compiling and preparing legal documents. It obtains judgments, statutes, and case briefs from valid sources like IndianKanoon, SCC Online, and e-Courts. As primary legal texts hold much inconsistency, noise, and irrelevant portions (such as page numbers, headers, or scanned OCR noise), the preprocessing step makes them uniform and structure-bound. Critical steps involved are eliminating of redundant text, normalization of quotation formats, anonymization of sensitive details, and normalization of terminologies across cases. In addition, the module makes sentence segmentation, tokenization, and named-entity recognition to capture legal entities such as names of parties, sections of the IPC, provisions of the Constitution, and precedents being cited. By doing so, this module makes sure that downstream AI/ML models receive tidy, machine-readable, and contextually rich inputs.

2. Knowledge Representation Module

Legal documents are necessarily complex, consisting of several layers of facts, statutes, arguments, and judicial reasoning. Aiming at capturing such nuances, this module produces a structured knowledge representation in the form of graphs. A case is mapped into a law–fact graph, whose nodes are statutes, charges, facts, and precedents, and whose edges are their semantic and legal relationships (e.g., “support a charge” or “apply a precedent” to a statute). The graph is augmented with extra-knowledge such as statutory definitions, cross-references, and historical interpretations, which allows the system to reason out of plain text. The built knowledge base is indexed in a vector database or knowledge graph engine for high-speed retrieval and reasoning support. This module occupies a pivotal position between bridging unstructured text data and structured legal reasoning.

3. Semantic Embedding Module

While search based on keyword takes care of similarity at the surface level, legal reasoning needs a much deeper insight into the world of semantics. This module takes care of that challenge through conversion of judgments, statute, and legal queries into dense vector embeddings. These embeddings are created using domain-specific language models like InLegalBERT, Legal-LLaMA, or fine-tuned transformer models trained on Indian legal databases. These embeddings are created at several granularities: sentence-level (to incorporate facts on a fine-grained level), paragraph-level (to incorporate legal argumentation), and case-level (to incorporate gross reasoning of judgment). Through this, the system maintains multi-granularity

understanding of legal text, allowing both fine-grained match (e.g., identifying a statute corresponding to a specific sentence) and coarse-grained retrieval (e.g., retrieval of a complete case resembling the current facts). This module, therefore, supplies the semantic backbone for all retrieval and reasoning activities.

4. Retrieval & Similar Case Finder Module

Precedents make up the essence of judicial decision-making decisions in the legal field. This module shall fetch the best relevant cases and statutes for a specific fact description or legal question. Through a retrieval-augmented generation (RAG) pipe, queries are embedded and matched against case and statute stored vector representations using FAISS or Elasticsearch tools. The system returns the top-k most relevant documents, ranked according to a mix of semantic similarity, frequency of citation, outcome similarity, and time relevance. For instance, when presented with a criminal case based on circumstantial evidence for murder, the system is able to return the most frequently cited precedents for IPC 302 and highlight the reasoning that has been applied. This allows judges, lawyers, and researchers at a click to have access to the top legally relevant precedents without being forced to ruffle through thousand documents, manually.

5. Multi-Level Attention & Context Integration Module

This is the system's core reasoning engine, tasked with collating facts, statute, and precedents within a common decision space. Employing sophisticated attention mechanisms, the module mimics the practice of human judges balancing evidence and law. Cross-attention layers connect case facts with statute, so descriptions of facts are matched with the legal provisions on which they rely. Graph attention mechanisms reason over the law–fact graph, giving special prominence to key nodes like oft-cited statute or determinative precedent. Hierarchical attention permits information at many different levels—individual sentence, judgment, or collection of precedents—to be aggregated, yet remain relevant within context. By combining multi-source information thus, the system builds a contextually aware grasp of legal problems, which allows for both accurate prediction and explainable reasoning.

6. Legal Judgment Assistance & Drafting Module

This module maps the system's reasoning into action-able legal outputs for the user. It creates structured case summaries, such as facts, issues framed, relevant statutes, arguments taken, and ultimate decisions. For new questions, it is also capable of predicting likely charges, relevant sections of the IPC, and even ranges of sentences based on case history. Apart from prediction, the module even helps create draft judgments. Through generative language models trained on judicial writing style, it

creates draft orders conforming to the typical courtroom structure—introduction, facts, issues, reasoning, statutes relied on, precedents relied on, and final ruling. These drafts can be had by judges or lawyers at a starting point, conserving time and maintaining uniformity in legal documents.

7. User Interaction Module

A high-end backend is worth little unless end-users find it easy to interact with brilliantly. This module gives both lawyers/lawyers-to-be and laypersons an interface for using the system. The interface accommodates natural language queries (e.g., “Find cases under IPC Section 302 involving circumstantial evidence”) as much as structured input forms. The output is user specific: a lawyer/judge might get detailed precedent graphs and statute linkages, but a citizen might get plain-English explanations of charges/orders-in-their-favour, with simple outcome explanations. Other features incorporate search filters by year, court, or statute, and visualization modules for presenting legal graphs and attention heatmaps. This allows for accessibility, usability, and widespread adoption of the system.

8. Evaluation & Explainability Module

With the sensitivity of legal decision-making, reliability and transparency are crucial. This module is committed to performance evaluation of the system and guaranteeing explainability. The models are aliquot using metrics like recall, precision, F1-score (for predicting statute), and ROUGE/BLEU (for judgment text synthesis). Besides quantitative measures, the module avails interpretability using attention heatmaps which reveal which facts, statutes, and statutes affected a prediction. Users trace the system's deliberations towards its outputs, gaining confidence in its recommendations. Human-in-the-loop evaluation is also performed, whereby legal experts confirm the correct retrievals of precedents, correct statute predictions, and synthesized draft judgments. This guarantees that the system does not only yield correct results but also transparent and defensible reasoning.

9. Cybersecurity & Data Protection Module

Legal documents hold highly sensitive personal and institutional information, particularly when dealing with cases of cybercrime, financial scams, or company disputes. This module takes utmost care for the highest level of cybersecurity, using encryption mechanisms for the protection of storing and transmitting case files, using access controls for varying user groups (such as judges, lawyers, researchers, or civilians), hashing for ensuring data integrity, and special care for anonymizing personally identifiable information (PII) and sensitive digital evidence like IP logs,

transaction histories, or intercepted communications, which are rampant for cybersecurity cases. Apart from this, there is also the provision for adding special modules for identifying anomalies within digital evidence, authenticating electronic records under provisions such as Section 65B of the Indian Evidence Act, and for tracking chain-of-custody tracking. This not only makes the system impregnable for attacks on the Cyber Front, but it also makes it adaptable for legal research for areas where cybersecurity plays the main role.

10.High-Performance Computing & Scalability Module

Certain case classes, for instance, large financial scams, company insolvency disputes, or mass cases involving a thousand or more documents, require high computational power. This module gives the system the capability of managing such high-computation workloads at full efficiency. Distributed computing architectures, GPU acceleration, and cloud-elastic resources are employed for scalability purposes. Large data sets are computed in parallel using Apache Spark or Ray, for instance, while deep learning workloads are spread across GPUs/TPUs for quick training, as well as quick inference. The system also uses model-parallelism, knowledge-distillation, techniques for performance, and balancing efficiency, respectively. Incorporating the module makes the assistant not just suitable for day-to-day case law research but for computationally intensive legal analysis jobs such as multi-document summarizations, bulk precedents retrieval, or executing predictive models on a thousand concurrent cases, respectively. This makes the system remain dependable, responsive, yet future-proof even at high-demand usage, respectively.

11.Evidence Management & Digital Validation Module

This module deals with case management of exhibits and evidence presented at trials. It facilitates ingestion of electronic evidence like audio, video, and forensic reports, and verifies them through hash-checking and metadata validation. It is also capable of cross-verifying if the evidence adheres to the provisions of Section 65B of the Indian Evidence Act for electronic records. Judges are benefited with sound evidence tracking, authenticity verification on a mechanical basis, and a standardized method of referring to exhibits while giving judgments.

12.Courtroom Workflow Integration Module

This module is built for easy integration with already implemented court management modules (such as e-Courts, Case Information System, or NJDG). Judges can ingest case files directly, monitor hearings, and send AI-enabled drafts directly into their case workflow. It minimizes manual steps and makes sure that the assistant augments installed judicial IT infrastructure.

13.Real-Time Query & Interactive Analytics Module

This permits judges to retrieve information from the system on a dynamic basis during hearings. For instance, when a lawyer references a precedent at a hearing, the judge might immediately ask the assistant for “other cases when Section 307 IPC was recourse for similar occasions.” The module gives visual analytics (timelines, graphs, statistics) for comparison of precedents at a speedy pace. This functions like a real-time research assistant at hearings.

14.Multi-Language Legal Understanding Module

Because a large number of Indian cases consist of local languages, the module facilitates multi-lingual processing of legal documents. Through translation and cross-lingual embeddings, it is capable of translating case texts from Hindi, Tamil, Bengali, and other local languages into English statutes and precedents, matching them with them. Regional judges are beneficiaries since they are able to view relevant precedents regardless of the language of writing of the original precedents.

EXPECTED OUTCOME:

1. Improved Access to Legal Precedents

Judges will have immediate access to the most relevant precedents without the need for manual perusing of extensive case files. Through the application of semantic embeddings and retrieval-augmented search, the system will recognize top-cited judgments and connect them with relevant statutes for application. This effect minimizes judges' time on legal research, facilitates more standardized citation of previous decisions, and facilitates uniformity across the judgship.

2. Enhanced Accuracy in Statute and Charge Identification

The system will, on its own, highlight charges and relevant statutes according to the facts of a case. For example, if a given fact pattern points towards murder under IPC 302, or towards cheating under IPC 420, the system will actively fetch them and display them. This makes sure that judges never miss important provisions and fortifies the quality of judicial reasoning by inscribing it according to the proper statutory regime.

3. Efficient Evidence Management and Validation

Dedicated modules for digital evidence, automates verification on authenticity, compliance with Section 65B of the Evidence Act, and validation of metadata. This helps ensure that spurious, manipulated, or irrelevant evidence is raised at an early stage, minimizing disagreement at the courtroom level. Judges have confidence that the evidence being

referenced is sound and traceable, thus fortifying the purity of hearings.

4. Real-Time Decision Support in Courtrooms

The analytical module for interactivity will enable judges at hearings to interrogate the system. For instance, when a lawyer cites a precedent, a judge can at once confirm its context, identify similar cases, or consider counter-arguments on a real-time basis. This consequence makes the system a courtroom aide that extends a judge's capacity promptly to react swiftly and conclusively under changing arguments.

5. Transparent and Explainable AI Reasoning

Multiple attention levels guarantee that the predictions are transparent. Judges won't just observe what statute or precedent is being proposed, but also why it was proposed. Legal graphs and attention heatmaps will display visibly which facts impacted the prediction, revealing the AI as transparent. This is a trusted outcome, which makes the system an auxiliary, not a black-box decision-maker.

6. Faster Judgment Drafting and Documentation

Using the judgment drafting module, the system shall prepare structured drafts according to the traditional legal form (facts, issues, reasoning, statute applied, precedents referred, decision). Judges need only correct or fine-tune such drafts rather than prepare from the beginning. This result minimizes paperwork burden, facilitates fast case disposal, and makes decisions maintain Court consistency.

7. Strengthened Cybersecurity in Legal Data Handling

Judges will be aided with tight data protection, encryption, and case files access control. Personal and institutional sensitive data, particularly in financial fraud and cybercrime cases, will be secured against anyone tampering or accessing without authority. This ruling ensures judicial data, which is highly confidential, is dealt with the ultimate cybersecurity measures.

8. High Scalability for Complex and Bulk Cases

The system will seamlessly handle high-volume litigations, mass petitions, or documents running into thousand numbers using cloud-based distributed computing, and GPU acceleration. Judges for such high-computation matters will get timely information without being held up. This results in scalability from a simple single-case examination to large multi-party disputes.

9. Multi-Language Legal Accessibility

Regional court judges will have exposure to precedents and judgments over several Indian languages. A Tamil case, for instance, can be matched with Hindi or English precedents using cross-lingual embeddings. This consequence facilitates inclusivity, provides equal justice-justice coverage across regions, and transcends societal, judicial, or legal-linguistic barriers.

10. Reduction of Bias and Fairness Monitoring

By adding fairness tracking, the system will emphasize the recommended biases (e.g., biased precedent selection). Judges will have the confidence of being more balanced in decisions due to AI recommendations being verified for impartiality. This is a direct contribution to ethical AI implementation within the judiciary.

11. Integration with Existing Judicial Workflows

Integration with e-Courts and NJDG without seams, judges will be able to view AI-aided insights without ever exiting their current workflow systems. Draft orders, case summaries, and extracted precedents may be pushed back into court management applications. This is a usability enhancement ensuring easy adoption without interfering with established routines.

12. Increased Judicial Productivity and Cost Savings

By adding fairness tracking, the system will emphasize the recommended biases (e.g., biased precedent selection). Judges will have the confidence of being more balanced in decisions due to AI recommendations being verified for impartiality. This is a direct contribution to ethical AI implementation within the judiciary.

13. Improved Public Trust and Transparency

When decisions are supported with clear AI explanations, corroborated evidence, and precedents, citizens have greater confidence in the judicial branch. The branch makes certain that not just justice is done, but that it is apparant it is done through explained, organized outputs. This makes judicial decision-making more legitimate.

14. Empowerment of Judges through Decision Support

Rather than supplanting judicial deliberation, the system versifies judges by behaving as a legal research assistant. Judges are still sole arbiters, but with improved knowledge backup. This is a consequence that preserves judicial autonomy while increasing their capacity to

manage expanding volumes of business through confidence.

STATE-LEVEL BENEFITS

1. Reduction of Case Backlogs

One of the biggest challenges in state-level judiciary systems is the backlog of pending cases, which often stretches into years. By automating research, evidence validation, and judgment drafting, the system significantly reduces the time judges spend on each case. Faster processing directly translates into higher case disposal rates, helping state courts clear long-pending matters more effectively.

2. Consistency in Legal Decisions Across Districts

State courts often face issues of inconsistency where similar cases in different districts may result in varying interpretations of law. With this system providing standardized precedent retrieval and statute mapping, judicial reasoning becomes more uniform across the state. This strengthens fairness, prevents contradictory rulings, and promotes a unified interpretation of state and central laws.

3. Increased Judicial Productivity

By automating time-intensive tasks such as document review, statute identification, and precedent search, judges can focus on the core reasoning process. This leads to higher productivity across the judiciary in the state, enabling judges to handle more cases per year without compromising quality.

4. Enhanced Transparency and Public Trust

At the state level, public trust in the judiciary is critical. By providing explainable AI outputs, clear reasoning trails, and validated evidence, the system ensures that judgments are transparent. Citizens gain confidence in the judicial system when they see that outcomes are backed by fair, structured, and explainable processes.

5. Strengthened Cybersecurity for State Judicial Data

With rising cybercrime, ensuring the safety of judicial records is essential. The cybersecurity module protects state court records, evidence, and case files from tampering or unauthorized access. This prevents data breaches, ensures compliance with digital evidence laws, and guarantees the integrity of the judicial system at the state level.

6. Faster Resolution of Cybercrime and Digital Evidence Cases

State courts increasingly deal with cybercrime, digital fraud, and data-related offenses. The system's evidence validation and cybersecurity modules help judges evaluate electronic evidence quickly and confidently. This ensures faster resolution of cybercrime cases, which is crucial for states grappling with rising digital offenses.

7. Multi-Language Support for Regional Justice Delivery

State-level courts often handle cases in regional languages, creating challenges when consulting national precedents. With multilingual embeddings and translation support, this system ensures that judgments in Tamil, Hindi, Telugu, or Bengali can be connected with English or other regional precedents. This strengthens inclusivity and ensures that no linguistic barriers hinder justice delivery at the state level.

8. Better Coordination with State Legal Services Authority (SLSA)

The system can integrate with state-level legal services authorities to assist in legal aid cases. Judges dealing with underprivileged litigants can use AI support to ensure fair representation and access to the correct statutes and precedents. This enhances the effectiveness of state-level legal aid initiatives.

9. Cost Reduction in Judicial Administration

By digitizing research, retrieval, and drafting, the system reduces reliance on manual clerical staff for tasks like citation management and document review. This translates into administrative cost savings for the state judiciary, freeing up resources that can be redirected to infrastructure or legal aid.

10. Capacity Building and Training for Judges

The system can act as a knowledge hub for new judges at the state level, helping them quickly understand statutes, precedents, and case patterns. It serves as a continuous training tool, improving legal knowledge dissemination across the state judiciary. This strengthens institutional capacity and ensures better decision-making even among newly appointed judges.

11. Real-Time Data for State Judicial Policy

Aggregated usage data from the system (such as most-cited statutes, common types of disputes, or regions with heavy backlog) can provide valuable insights to state judicial policymakers. This allows High Courts and State Legal Commissions to identify problem

areas and design targeted reforms.

12. Contribution to State-Wide Digital Judiciary Initiatives

Many states are pushing for e-Courts, digitized records, and AI adoption in governance. This system integrates seamlessly with those efforts, showcasing the state judiciary as technologically advanced and future-ready. It also supports state-wide digital transformation goals in line with national initiatives like Digital India.

CONCLUSION

The AI-Driven Legal Document Research Assistant with Advanced Context Comprehension is a revolutionary solution that facilitates streamlined legal research, precedent extraction, statute tagging, and judgment writing. Through the incorporation of sophisticated modules like knowledge graphs, semantic embeddings, and multi-level attention, the system closes the gap between unstructured legal documents and structured reasoning, allowing judges to make quicker, more precise, and contextually aware decisions.

Besides main functions, the solution includes special modules for cybersecurity, high-performance computing, proof validation, and international accessibility, which enable it to be flexible enough for both every day and high-complexity cases. Through such features, the system not only accompanies legal experts on a daily basis but also enhances judicial transparency, justice, and effectiveness. At the state level, the solution directly helps decrease case backlog, make decisions steadily across courts, and enable judges with decision support at hearings on a real time basis.

Finally, the project is also in sync with the country-wide judicial modernization initiatives like Digital India and e-Courts, presenting itself as a progressive initiative increasing accessibility, scalability, and confidence in the judiciary. Instead of supplanting human intel, it is an intelligent aide, increasing judges' capability to concentrate on primary reasoning while managing sophisticated data and work-flows. That makes a wiser, quicker, and clearer justice system, providing timely and fair results for society.

Annexure - I

Format for Appraisal of the Department

(To be forwarded along with the detailed proposal for funding)

1	GENERAL INFORMATION	
1.1	Project Title	AI-Driven Legal Document Research Assistant with Enhanced Context Understanding
1.2	Name of the Secretariat Department	Department of Justice, Ministry of Law and Justice (Government of India)
1.3	Name of the implementing agency (Please refer para 2 (II) of G.O.(MS)No.69)	State Planning Commission (SPC), Tamil Nadu
1.4	Sector	Law, Justice and Governance Sector
1.5	Has this project or a similar project been appraised by the State Planning Commission in previous rounds from your Department? If yes, details to be annexed	No
2	PROBLEM STATEMENT	
2.1	<p>Describe the scope of the problem that is being addressed through this proposal.</p> <p><i>Summarize the problem with relevant details within 2-3 paragraphs.</i></p>	<p>The Indian judicial system is one of the largest in the world, but it is overburdened with an ever-increasing number of pending cases. According to the National Judicial Data Grid, crores of cases remain unresolved across High Courts and subordinate courts, often leading to delays that affect citizens' right to timely justice. A significant portion of judicial time is consumed in labor-intensive tasks such as searching for relevant statutes, reviewing large volumes of case law, validating evidence, and drafting detailed judgments. The absence of efficient technological support makes these processes slow, inconsistent, and highly resource-intensive, contributing to the growing backlog of cases.</p> <p>Judges and lawyers also face challenges in accessing relevant precedents and statutes quickly, especially when cases involve complex legal issues, digital evidence, or multilingual documentation. Manual research often leads to inconsistencies across judgments delivered in different courts, reducing predictability and fairness in the legal system. With the increasing volume of cybercrime, financial fraud, and large-scale litigation, the burden on judicial resources has grown beyond traditional methods of case handling.</p>

		<p>This proposal addresses these challenges by introducing an AI-driven legal research and judgment assistance system that leverages advanced technologies such as natural language processing, knowledge graphs, retrieval-augmented reasoning, and explainable AI. By streamlining research, ensuring transparency, validating evidence, and providing decision support, the system aims to empower judges, reduce case backlogs, and enhance the efficiency and trustworthiness of the judicial process at both state and national levels.</p>
2.2	<p>What is the geographic, organizational, and cultural context for the problem? Who does the problem impact and how does it impact them?</p>	<p>The problem exists across the Indian judicial landscape, encompassing the Supreme Court, High Courts, and subordinate courts in every state, including Tamil Nadu. Geographically, the impact is most severe at the state and district court levels, where the majority of cases are filed and where pendency is highest. Tamil Nadu alone has lakhs of pending cases in its High Court and district courts, reflecting a systemic challenge that spans urban, semi-urban, and rural regions. This wide geographic spread makes it difficult to ensure uniform access to precedents, statutes, and digital case resources without technological intervention.</p> <p>Organizationally, the issue impacts multiple stakeholders within the judicial ecosystem. Judges and magistrates face overwhelming workloads due to time-consuming research and drafting. Lawyers spend significant effort manually retrieving relevant precedents, often under strict time constraints, which affects the quality of arguments presented. Court administrative staff struggle with managing bulky records, verifying evidence, and digitizing files. At a governance level, state legal departments and judicial councils are under pressure to reduce pendency rates and improve disposal efficiency, but lack the right technological support to meet these demands.</p> <p>Culturally, the problem directly impacts citizens and litigants, who often face years of delays in receiving justice. Delayed case resolutions weaken public trust in the judiciary, particularly for vulnerable groups who lack the resources to sustain prolonged legal battles. Language diversity further complicates access, as many judgments and legal documents are recorded in English while proceedings occur in regional languages. This gap creates barriers for both citizens and lower courts in effectively accessing and interpreting relevant precedents. Collectively, the burden of delay, inconsistency, and limited accessibility undermines the principle of “justice delayed is justice denied,” making this a problem of critical societal importance.</p>

2.3	<p>Please share any data, studies or articles that you have referred to, to support/evidence this problem statement.</p> <p><i>This could also be data from your department's work at the grassroot level or from any published journals.</i></p>	<ol style="list-style-type: none"> 1. Anuj Kumar, Vishwas Raj, and V. S. Dixit, "NyayaAnumana: A Dataset for Legal Judgment Prediction in Indian Courts," <i>Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)</i>, pp. 1125–1137, 2023. 2. Shubham Nigam, Aayush Bhatt, and S. Saha, "InLegalBERT: A Pre-trained Language Model for Indian Legal Texts," <i>International Conference on Computational Linguistics and Legal Informatics (COLIEE)</i>, 2022. 3. M. Chalkidis, I. Androutsopoulos, and N. Aletras, "Legal Judgment Prediction: A Survey of Approaches and Open Challenges," <i>Artificial Intelligence Review</i>, vol. 55, no. 5, pp. 3301–3338, 2022. 4. T. Verma, S. D. Bhatt, and K. Sharma, "Estimating Time to Clear Pendency of Cases in High Courts in India," <i>arXiv preprint</i>, arXiv:2307.12549, 2023. 5. N. Gupta and R. S. Yadav, "Factors Affecting Efficient Discharge of Judicial Functions in India," <i>Social Science Research</i>, vol. 113, pp. 45–58, 2024. 6. R. Bhatnagar and N. Huchhanavar, "Predicting Delays in Indian Lower Courts Using AutoML and Decision Forests," <i>arXiv preprint</i>, arXiv:2307.16285, 2023. 7. S. Mishra, "Cyclic Syndrome of Arrears and Efficiency of Indian Judiciary," <i>Journal of Law and Society</i>, vol. 49, no. 4, pp. 721–742, 2022. 8. S. Krishnan and V. Kumar, "Delay in Process, Denial of Justice: Jurisprudence and Empirics of Criminal Trial Delays in India," <i>Indiana University Legal Studies Research Paper</i>, pp. 1–28, 2011. 9. Ministry of Law and Justice, Government of India, "National Judicial Data Grid (NJDG): Statistics on Case Pendency," <i>Department of Justice Reports</i>, 2025. 10. Supreme Court of India, "Annual Report on Judicial Backlog and Case Disposal," <i>Supreme Court Publications</i>, New Delhi, 2024.
2.4	<p>Can this problem be addressed through the regular department budgets?</p>	<p>Yes/No</p>
2.5	<p>If no, please describe why is this problem unique and requires an innovative solution</p> <p><i>You may wish to describe why this cannot be considered by the regular departmental budgets and requires budgets from Special Initiative. Is the</i></p>	<p>Traditional departmental allocations are designed to cover routine judicial expenditures such as salaries, infrastructure maintenance, administrative expenses, and incremental digitization efforts. While these budgets sustain the existing system, they do not have the flexibility or scope to fund advanced, research-driven, and high-computation technologies such as artificial intelligence, knowledge graphs, retrieval-augmented reasoning, or multilingual embeddings.</p> <p>This problem is unique because it directly addresses the systemic backlog, inconsistency, and inefficiency in the judicial process—issues that have persisted for decades and cannot be solved through incremental resource allocation. Regular budgets cover day-to-day court functioning, but they cannot provide for the</p>

	<p><i>problem being addressed important enough to warrant an ‘innovation’? (for instance, identification of a gap in the existing services or an issue not tackled so far, but needs to be addressed)</i></p>	<p>development of innovative, state-of-the-art tools that require cross-disciplinary research, specialized AI expertise, high-performance computing infrastructure, and integration with existing e-Courts platforms. Such requirements fall outside the scope of routine financial provisioning.</p> <p>The importance of this problem warrants classification under a Special Innovation Initiative because it fills a critical gap in existing services. Current judicial IT systems largely focus on digitization and case management, but they do not provide intelligent decision-support features such as statute prediction, precedent retrieval, evidence validation, or real-time analytics for judges. Without innovation-focused funding, the judiciary risks continuing with outdated processes that are incapable of addressing the scale and complexity of modern legal challenges, particularly in cybercrime, financial fraud, and multi-party litigation. By supporting this project as an innovation, the state can pioneer a model that not only improves judicial efficiency but also sets a national benchmark in applying AI for governance and justice delivery.</p>
2.6	<p>What is the source of the proposed solution?</p>	<p>The proposed solution originates from a combination of academic research, technological advancements, and identified gaps in the existing judicial system. It draws upon recent studies in the field of legal artificial intelligence, including models such as InLegalBERT and NyayaAnumana datasets for Indian judgments, as well as global research on legal judgment prediction, retrieval-augmented generation (RAG), and graph-based reasoning. These provide a strong scientific foundation for applying natural language processing and machine learning to the Indian legal context.</p> <p>At the practical level, the solution is informed by ground realities of the Indian judiciary, as documented by the National Judicial Data Grid (NJDG), Supreme Court annual reports, and empirical studies highlighting massive pendency, delays, and inefficiencies in accessing statutes and precedents. By combining these insights with innovations from AI research and high-performance computing, the solution is designed to go beyond routine digitization and provide intelligent decision support to judges.</p> <p>Thus, the source of the solution lies in the intersection of cutting-edge AI research and the pressing judicial need for efficiency, transparency, and consistency, making it both academically grounded and socially relevant.</p>
2.7	<p>Are there any alternative solutions for the proposed challenge? If yes, has the department explored them?</p> <p><i>Please explain examples of different ways of approaching the problem that you may have explored before arriving at the proposed solution.</i></p>	<p>Yes, there are alternative approaches to addressing the challenges of judicial backlog and inefficiency, but these have significant limitations compared to the proposed AI-driven solution.</p> <p>One alternative is increasing manpower by appointing more judges, clerks, and administrative staff. While this can reduce some burden, it is not sustainable in the long run due to budgetary constraints, recruitment delays, and the sheer scale of pendency (crores of cases). Moreover, additional manpower does not guarantee consistency or speed in legal research and precedent retrieval.</p>

3	PROJECT DETAILS	
3.1	<p>Describe the innovation solution that is being proposed to address this challenge</p> <p><i>Not more than 150 words.</i></p>	<p>The proposed solution is an AI-Driven Legal Document Research Assistant with Enhanced Context Understanding that leverages artificial intelligence, natural language processing, and graph-based reasoning to support judges in research and decision-making. Unlike existing digitization or keyword-based search tools, the system uses domain-tuned language models (InLegalBERT, Legal-LLaMA) to generate semantic embeddings, a knowledge graph to map statutes, facts, and precedents, and retrieval-augmented reasoning to provide judges with the most relevant cases and statutes. Multi-level attention mechanisms integrate facts with applicable laws to deliver contextually accurate insights.</p> <p>The system further incorporates modules for cybersecurity, evidence validation, high-performance computing, multilingual support, and real-time courtroom analytics, making it robust and future-ready. Judges benefit through faster precedent retrieval, transparent and explainable outputs, and AI-assisted judgment drafting, reducing backlog and improving consistency. This solution represents a transformative innovation that goes beyond routine digitization to empower the judiciary with smart, transparent, and scalable decision support.</p>
3.2	<p>Please select the type of innovation involved in your project</p>	<p>New/Novel/Unconventional I/Creative Intervention/Approach/ Organizational Model/ Operational Process/Service to tackle the identified issue/ problem that would lead to a substantial improvement in developmental/governance outcomes</p>
3.3	<p>Please share any successful case studies of similar innovation that you may have referenced.</p>	<p>CaseLaw Access Project (Harvard Law School, USA) Harvard digitized over 40 million pages of U.S. case law and applied AI-driven search and retrieval tools to make judgments accessible to lawyers, researchers, and citizens. This project demonstrated how large-scale digitization combined with intelligent search can transform legal research.</p> <p>LexisNexis and Westlaw (Global Legal Tech Platforms) These platforms use advanced search, natural language processing, and citation mapping to assist lawyers in retrieving statutes and precedents. While commercial, they serve as proof that AI-driven research assistants can dramatically reduce the time required for legal research and improve accuracy.</p> <p>COLIEE (Competition on Legal Information Extraction and Entailment, Japan/Canada) Academic competitions like COLIEE benchmark AI models for statute retrieval, case entailment, and legal reasoning. Successful systems from COLIEE have proven the feasibility of applying NLP and machine learning to automate complex legal reasoning</p>

		<p>tasks.</p> <p>NyayaAnumana Project (India, 2023) A recent academic initiative that released a large-scale dataset of Indian legal judgments and introduced baseline AI models for outcome prediction. Though focused on judgment prediction, it validates the applicability of AI to Indian judicial data.</p> <p>Estonia's e-Court and AI Pilot (Europe) Estonia experimented with AI systems to help small claims courts draft preliminary decisions. This case study illustrates how AI can be responsibly integrated into judicial workflows without replacing judicial discretion.</p>
3.4	<p>Does the proposal include a project implementation plan with a time schedule. If yes, please share it as an annexure to the proposal.</p>	<p>Yes/No</p> <p>Yes (Annexure III – Time Schedule Plan) - Enclosed</p>
3.5	<p>If no, please share the tentative date by when this will be submitted</p>	---
4	PROJECT OUTCOMES	
4.1	<p>Describe the economic benefits proposed through the implementation of this project <i>Not more than 100 words</i></p>	<p>The implementation of this project will reduce the enormous economic costs associated with prolonged litigation, repeated hearings, and delayed case disposal. By enabling faster precedent retrieval, automated evidence validation, and AI-assisted judgment drafting, the judiciary can clear backlogs more efficiently, lowering operational and administrative expenses. Citizens and businesses also benefit from reduced legal expenditure due to shorter trial durations and quicker resolutions. At the state level, this leads to significant savings in court infrastructure and manpower costs, while fostering a more predictable legal environment that supports investment, business growth, and overall economic development.</p>
4.2	<p>Is the project scalable through your department on successful implementation of the pilot?</p>	Yes

4.3	If yes, indicate the department's plans	<ul style="list-style-type: none"> Expand the use of AI-driven legal technologies, multilingual interfaces, and real-time analytics across district and High Courts after the pilot phase. Integrate knowledge graphs, retrieval-augmented reasoning, and multi-level attention models for consistent and efficient legal research at scale. Fine-tune AI algorithms for statute prediction, precedent retrieval, evidence validation, and judgment drafting across diverse case categories (civil, criminal, cybercrime, financial disputes). Deploy cloud-based platforms for secure data storage, scalable computing, and seamless integration with e-Courts and NJDG systems. Introduce voice-enabled and multilingual interfaces to ensure accessibility for judges, lawyers, and staff with varying digital literacy. Provide continuous technical support, judicial training programs, and regular system upgrades to ensure smooth adoption and long-term sustainability.
4.4	Does the project include continued financial sustainability plans?	<p>Yes/No</p> <p>Yes, the project includes continued financial sustainability plans.</p>
4.5	If yes, indicate the departments plans	<p>The Department Plans are:</p> <ul style="list-style-type: none"> Establish subscription-based or pay-per-use service models for courts, legal institutions, and law firms to support ongoing system maintenance and upgrades. Explore partnerships with private legal-tech firms, bar associations, and academic institutions to share costs and expand access. Leverage government schemes and subsidies under Digital Governance and e-Courts initiatives to reduce reliance on one-time funding. Implement capacity-building programs for judges, lawyers, and court staff to ensure local ownership and long-term management. Commercialize data-driven legal analytics and decision-support services to attract private investors and generate continuous revenue. Ensure regular system upgrades, technical support, and innovation cycles funded through the above sustainable revenue streams.
5	FINANCIAL	
5.1	Please mention the total project budget requirement in rupees.	₹ 50,00,000 (Rupees Fifty Lakh Only)
5.2	Please indicate the project duration	3 Years
5.3	Other sources of funding	Institutional support - using the existing facilities for the development

	if available	of the project
5.4	Cost of the machinery included in the project	<p>High-Performance Servers / GPU Workstations – For training and deploying AI/ML models (₹ 5 Lakhs)</p> <p>Cloud Infrastructure Setup – Hybrid cloud/on-premise storage, distributed compute clusters (₹ 5 Lakhs)</p> <p>Secure Storage Systems – Encrypted storage servers for legal documents and case files (₹ 3 Lakhs)</p> <p>Networking & Security Appliances – Firewalls, VPNs, and intrusion detection systems to ensure cybersecurity (₹ 2 Lakhs)</p> <p>Backup & Disaster Recovery Systems – High-capacity storage devices and redundancy hardware (₹ 1 Lakhs)</p> <p>User Interface Hardware – Courtroom terminals, secured tablets, or kiosks for judge/staff access (₹ 1 Lakhs)</p>
5.5	Percentage of the Machinery cost to the project cost	35 %
5.6	Cost of the Computers and Accessories included in the project	20 Lakhs
5.7	Percentage of the Computers and Accessories included in the project	40 %

5.8	Is the purchase of new machinery indispensable for Implementation of the project?	<p>Yes/No</p> <p>Yes</p> <p>The purchase of new machinery is indispensable for the effective implementation of the “AI-Driven Legal Document Research Assistant with Enhanced Context Understanding” project. This project is computation-intensive and relies heavily on advanced hardware to process large volumes of legal case documents, build graph-based reasoning models, and run AI algorithms for statute, charge, and sentencing predictions.</p> <p>Key components such as high-performance servers, GPUs, and scalable storage systems are essential for training and fine-tuning large legal language models. Similarly, dedicated computers and accessories are required for preprocessing millions of case files, running retrieval-augmented generation (RAG) pipelines, and enabling multi-attention reasoning in real time. Without modern computing infrastructure, the project’s objectives of accurate, explainable, and efficient legal judgment prediction cannot be achieved.</p>
5.9	Cost of experts / consultants included in the project cost	2.5 lakhs
5.1 1	Percentage of the Cost of experts / consultants to the project cost	5 %
5.1 2	Does the proposal include break up of budgets as Capital Expenditure and Revenue Expenditure?	<p>Yes,</p> <p>Yes, the proposal provides a clear separation into Capital Expenditure (CapEx) and Revenue Expenditure (RevEx) to ensure smooth fund sanctioning and compliance with TANII’s financial protocols.</p> <p>Capital Expenditure (CapEx): ₹31,75,000 Revenue Expenditure (RevEx): ₹18,20,000 Total: ₹50,00,000</p>
6	Others	
6.1	Does the project require Statutory clearances	<p>Yes / No</p> <p>(if ‘Yes’, the details should be furnished)</p> <p>No</p>
6.2	Statutory clearances obtained	<p>Yes / No</p> <p>(if ‘No’, the reasons to be furnished)</p> <p>No</p>
6.3	Assessment on possible risks and challenges	<ul style="list-style-type: none"> • Data Sensitivity & Privacy – Court judgments and case records contain confidential details; compliance with the Digital Personal Data Protection Act, 2023, is essential. • Accuracy & Reliability of Predictions – Misclassification of statutes, charges, or sentencing could undermine trust; requires expert validation and continuous model fine-tuning. • High Computational Demands – Training and deploying large legal

		<p>language and graph models need advanced servers, GPUs, and scalable infrastructure.</p> <ul style="list-style-type: none"> ● Adoption Resistance – Judges, lawyers, and clerks may be cautious about relying on AI-driven insights; explainability and trust-building are critical. ● Integration Challenges – Linking with e-courts platforms and existing judicial information systems may require technical customization and policy coordination. ● Regulatory & Ethical Risks – Use of AI in legal decision support may face skepticism and require strict adherence to judicial guidelines. ● Skill Gap – Specialized expertise will be needed for system maintenance, updates, and interpretation of outputs. ● Sustainability & Funding – Long-term operation depends on consistent financial support or a viable revenue model.
6.4	Does the project have linkages with similar projects?	<p>Yes / No</p> <p>Yes, the project aligns with several national and institutional initiatives that focus on technology-driven legal research and justice delivery.</p> <ul style="list-style-type: none"> ● The Supreme Court's e-Courts Mission Mode Project promotes digitization of case records and ICT adoption in judiciary, which complements this project's AI-based legal reasoning. ● The National Judicial Data Grid (NJDG) provides structured case data, forming a foundation for analytics and AI integration. ● Academic initiatives such as NyayaAnumana and CJPE (Court Judgment Prediction with Explanation) projects explore judgment prediction and legal NLP, which align closely with this project's goals. ● Internationally, projects like COLIEE (Competition on Legal Information Extraction and Entailment) also focus on statute retrieval and case law reasoning, showing strong synergy with the proposed work.
6.5	Further steps need to complete project preparation	<p>To complete the project preparation for “AI-Driven Legal Document Research Assistant with Enhanced Context Understanding,” a feasibility study is needed to assess the availability and structure of legal datasets and customize AI models for Indian judicial documents. A pilot court dataset (e.g., High Court or Supreme Court cases) needs to be selected, and a stakeholder engagement strategy should be developed to ensure awareness and adoption among judges, lawyers, and legal researchers.</p>
6.6	Resources required to complete the project preparation process (Please detail: need to engage consultants including transaction advisors)	<p>To complete the project preparation process, the resources required are:</p> <ul style="list-style-type: none"> ● Engagement of domain experts and consultants, including legal scholars, judicial data specialists, and AI/ML experts. ● Transaction advisors may be needed to assist in designing sustainable funding and scaling models beyond the initial phase. ● Collaboration with IT service providers and software developers will be necessary to build AI-driven legal research platforms, document parsers, and user-friendly interfaces. ● Infrastructure support, such as access to high-performance

		<p>computing (HPC) clusters, secure cloud storage, and GPU resources, will be required to train and deploy large legal language models.</p> <ul style="list-style-type: none"> ● Resources for stakeholder consultations, workshops, and training sessions with judges, lawyers, and law students will be needed to ensure smooth adoption and effective usage of the system.
6.7	Role on responsibilities of involved parties	<p>Roles and Responsibilities of Involved Parties</p> <ul style="list-style-type: none"> ● Project Implementing Agency (Lead Institution / University) <ul style="list-style-type: none"> ○ Overall project management and coordination. ○ Data collection, preprocessing, and secure storage of legal documents. ○ Development and integration of AI/ML models with document research workflows. ● Domain Experts (Legal Scholars, Judicial Experts) <ul style="list-style-type: none"> ○ Provide guidance on legal interpretation, statutes, and judicial processes. ○ Validate the accuracy, relevance, and interpretability of AI-generated outputs. ○ Assist in designing evaluation benchmarks and use-case scenarios. ● Technical Team (AI/ML Engineers, Software Developers) <ul style="list-style-type: none"> ○ Build, train, and optimize NLP models for legal text understanding. ○ Develop graph-augmented reasoning and context-aware retrieval systems. ○ Create user interfaces and tools for researchers, judges, and lawyers. ● Transaction Advisors / Funding Consultants <ul style="list-style-type: none"> ○ Design sustainable financial models for scaling and long-term maintenance. ○ Explore partnerships with judiciary bodies, law firms, and government agencies. ● IT Infrastructure Providers (Cloud / HPC vendors) <ul style="list-style-type: none"> ○ Provision high-performance computing, GPU clusters, and secure cloud storage. ○ Ensure uptime, scalability, and cybersecurity of deployed platforms. ● Stakeholders (Judges, Lawyers, Law Students, Researchers) <ul style="list-style-type: none"> ○ Act as end-users to test and validate the system in real-world contexts. ○ Provide feedback for iterative improvements. ○ Facilitate wider adoption across the legal ecosystem.

Annexure II

TABLE 1: Overall Budget Requirement (Rs in lakhs)

Head	Budget requirement under TANII	Support from any other agency	Total
Capital Expenditure			
Infrastructure development	15,45,455		15,45,455
Sector Specific/Area of focus related machinery or equipment	9,09,091		9,09,091
Computers, Electronics and other accessories	7,27,273		7,27,273
Revenue Expenditure			
Software and other digital assets	4,54,545		4,54,545
Manpower	10,00,000		10,00,000
Travel	1,81,818		1,81,818
Other cost (specify)	1,81,818		1,81,818
Total	50,00,000	0	50,00,000

**Please add/modify lines items (only horizontal) as per requirements of your proposal*

**No table should be left blank, unless there is no line item of expenditure in the said table*

TABLE II: Year wise Budget Requirement (Rs in lakhs)

Head	1st YEAR			2nd YEAR			3rd YEAR		
	Budget requirement under TANII	Support from any other agency	Total	Budget requirement under TANII	Support from any other agency	Total	Budget requirement under TANII	Support from any other agency	Total
Capital Expenditure									
Infrastructure development	7,27,273			5,45,455			2,72,727		15,45,455
Sector Specific/Area of focus related machinery or equipment	4,54,545			2,72,727			1,81,818		9,09,091
Computers, Electronics and other accessories	3,63,636			2,72,727			90,909		7,27,273
Revenue Expenditure									
Software and other digital assets	1,81,818			1,45,455			1,27,273		4,54,545
Manpower	3,63,636			3,63,636			2,72,727		10,00,000
Travel	72,727			54,545			54,545		1,81,818
Other cost (specify) – Training, Consumables, and Maintenance	90,909			54,545			36,364		1,81,818
Total	22,63,636	0	0	17,09,090	0	0	9,36,363	0	50,00,000

**Please add/modify lines items (only horizontal) as per requirements of your proposal*

**No table should be left blank, unless there is no line item of expenditure in the said table*

**Estimates to be enclosed as per tender transparency act*

TABLE III: Capital Expenditure

CAPITAL EXPENDITURE						
S.No	Item description	Specification	Reference Schedule Rate (PWD, SSR, FSR, ELCOT, etc.)	Rate Per Unit	Quantity	Total Cost
1	Infrastructure Development	Mid-range GPU Server (2 × NVIDIA L40S / A30 equivalent, 256 GB RAM, 4 TB SSD)		7,75,000	2	15,50,000
2	Sector Specific/Area of focus related machinery or equipment	1 × High-Speed Document Scanner (100 ppm, OCR enabled) + 1 × Legal Data Processing Server (Dual Xeon, 128 GB RAM)		4,50,000	2 sets	9,00,000
3	Computers, Electronics and other accessories	Developer Workstations (Ryzen 9/Intel i9, 64 GB RAM, 1 TB SSD)		2,25,000	2	4,50,000
		Networking & A/V Equipment (NAS 20 TB, Router, Camera/Mic kit)		2,75,000	1 set	2,75,000
Total						31,75,000

**Please add/modify lines items (only horizontal) as per requirements of your proposal*

**No table should be left blank, unless there is no line item of expenditure in the said table*

**Estimates to be enclosed as per tender transparency act*

TABLE IV: Revenue Expenditure

REVENUE EXPENDITURE						
S.No	Item description	Specification	Reference Schedule Rate (TNeGA, ELCOT, etc.)	Rate Per Unit	Quantity	Total Cost
1	Software and other digital assets	Open-source + subsidized AI/ML tools (HuggingFace Pro, CaseMine/Manupatra access, NIC Cloud credits)		1,50,000	3 bundles/year	4,50,000
2	Manpower	Project Associate 1 Project Assistant		38000 26000	3,35,000 (INR / year)	10,00,000
3	Travel	Court Visits / Data Collection Training & Awareness Workshops National Conferences / Seminars		40,000 15,000 5,000	60,000 (INR / year)	1,80,000
4	Other cost (specify) Training, Consumables, and Maintenance	Workshops Data Acquisition Connectivity Equipment maintenance Consumables		25,000 20,000 10,000 8000	63,350 (INR / year)	1,90,000
TOTAL						18,20,000

**Please add/modify lines items (only horizontal) as per requirements of your proposal*

**No table should be left blank, unless there is no line item of expenditure in the said table*

**Estimates to be enclosed as per tender transparency act*

**For Manpower expenditure please include details on purpose and period of engagement.*

Annexure III (Time Schedule)

AI-Driven Legal Document Research Assistant with Enhanced Context Understanding

[illegible]