State Planning Commission
Government of Tamil Nadu

Tamil Nadu Innovation Initiative Scheme (TANII)
(2025-2026)

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 project AI-Driven Legal Document Research Assistant with Enhanced Context Understanding seeks to tackle the challenge of legal information overload by providing judges and lawyers with an intelligent, context-aware assistant that accelerates research and drafting without replacing judicial discretion. Traditional keyword searches often miss nuanced legal semantics, but this system leverages retrieval-augmented generation, multi-level attention, and graph-based reasoning to connect facts, statutes, and precedents more effectively. It offers automated case summarization, precedent and statute retrieval, draft order generation, and explainable reasoning, all within a judge-in-the-loop framework that ensures transparency and control. With modular toolkits tailored to specific domains like criminal law, cybercrime, financial offenses, constitutional matters, and environmental cases, the assistant adapts to varied judicial needs. By reducing research time, promoting consistency in decision-making, and supporting structured, citation-ready outputs, the project ultimately aims to enhance judicial efficiency, improve access to relevant knowledge, and contribute toward easing the backlog of cases in the legal 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 integrates the following key components:

The project integrates the following key components to create a robust, judge-in-the-loop legal research assistant. At its foundation, it employs advanced AI and NLP models such as InLegalBERT, Longformer, and fine-tuned transformers to understand the nuanced semantics of legal language, enabling the system to capture context beyond surface-level keywords. These models are paired with retrieval systems like FAISS, Milvus, and Elasticsearch to ensure that the most relevant statutes and precedents are surfaced quickly and accurately. Complementing this is a knowledge graph framework built with tools such as Neo4j and PyTorch Geometric, which maps the complex relationships between facts, laws, parties, and evidence, making it easier for judges and lawyers to navigate multi-layered cases. To strengthen reasoning, the project uses multi-level attention mechanisms, including cross-attention and graph attention networks, which align case facts with statutory provisions and precedents while preserving contextual depth.

On the practical side, the project integrates document processing pipelines using PyMuPDF, pdfplumber, and Tesseract OCR to extract structured information from case files, judgments, and supporting evidence. The backend is designed with FastAPI and Flask for machine learning services, supported by Node.js and Express for orchestration, while databases like PostgreSQL and MongoDB manage structured and unstructured legal data. For the user-facing judge console, technologies like React, TypeScript, and Streamlit provide an interactive environment with modules for summaries, precedent exploration, draft order generation, and explainability panels. The system is deployed securely using Docker, Kubernetes, and MLflow, with observability handled through Prometheus and Grafana, and compliance ensured via role-based access control, audit trails, and encryption. Explainability is powered by SHAP, Captum, and retrieval-linked rationales, ensuring the assistant works transparently as a support tool for judges rather than replacing judicial discretion.

Goals & Objectives:

Goals

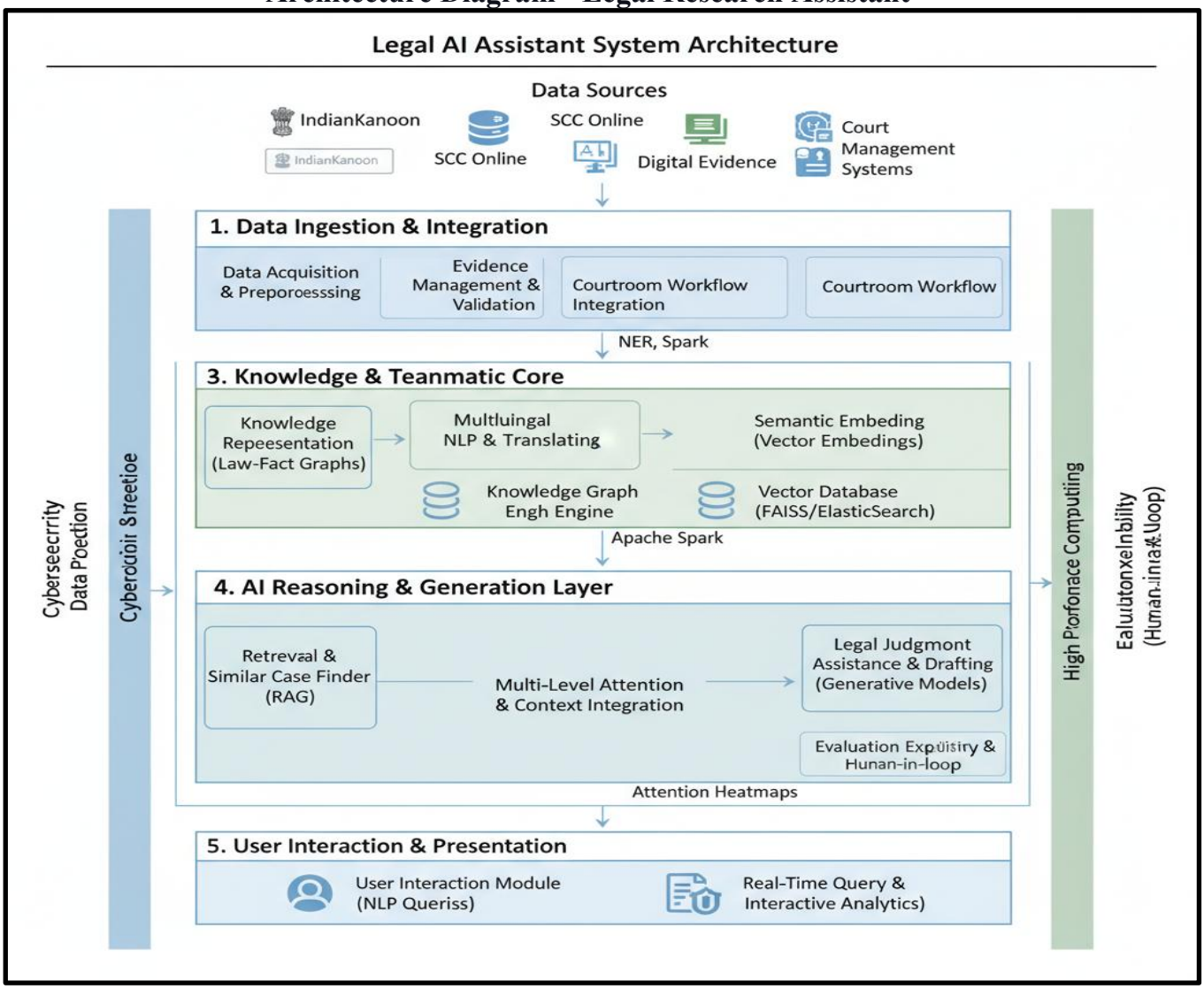
- Develop a **judge-in-the-loop AI assistant** that accelerates legal research and drafting without replacing judicial discretion.
- Reduce the **time and effort** required by judges and lawyers to analyze lengthy case documents, statutes, and precedents.
- Enhance **consistency and fairness** in judicial 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 is responsible for building the foundation of the system by gathering and preparing legal documents. It collects judgments, statutes, and case briefs from authentic sources such as IndianKanoon, SCC Online, and e-Courts. Since raw legal texts often contain inconsistencies, noise, and irrelevant sections (like page numbers, headers, or scanned OCR errors), the preprocessing step ensures uniformity and structure. Important tasks include removal of redundant text, normalization of citation formats, anonymization of sensitive information, and standardization of terminologies across cases. Further, the module applies sentence segmentation, tokenization, and named-entity recognition to identify legal entities such as party names, IPC sections, constitutional provisions, and cited precedents. By performing these tasks, this module ensures that the downstream AI/ML models receive clean, machine-readable, and contextually rich inputs.

2. Knowledge Representation Module

Legal documents are inherently complex, with multiple layers of facts, statutes, arguments, and judicial reasoning. To capture these nuances, this module creates a structured representation of knowledge in the form of graphs. Each case is transformed into a law–fact graph, where nodes represent statutes, charges, facts, and precedents, while edges represent their semantic and legal relationships (e.g., "fact supports charge" or "precedent applies to statute"). The graph is enriched with external knowledge such as statutory definitions, cross-references, and historical interpretations, enabling the system to reason beyond plain text. The constructed knowledge base is stored in a vector database or knowledge graph engine to support high-speed retrieval and reasoning. This module plays a critical role in bridging unstructured textual data with structured legal reasoning.

3. Semantic Embedding Module

While keyword-based search captures surface-level similarity, legal reasoning requires a much deeper understanding of semantics. This module addresses that challenge by converting judgments, statutes, and legal queries into dense vector embeddings. These embeddings are generated using domain-specific language models such as InLegalBERT, Legal-LLaMA, or fine-tuned transformer models trained on Indian legal corpora. The embeddings are produced at multiple levels: sentence-level (to capture factual details), paragraph-level (to capture legal arguments), and case-level (to capture overall judgment reasoning). By doing so, the system ensures multi-granularity understanding of legal texts, enabling both fine-

grained matching (e.g., identifying a statute related to a particular sentence) and coarse-grained retrieval (e.g., retrieving an entire case similar to the current facts). This module thus provides the semantic backbone for all retrieval and reasoning tasks.

4. Retrieval & Similar Case Finder Module

In the legal domain, precedents form the backbone of judicial decision-making. This module is designed to retrieve the most relevant cases and statutes for a given fact description or legal query. Using a retrieval-augmented generation (RAG) pipeline, queries are embedded and compared against stored vector representations of cases and statutes using tools such as FAISS or ElasticSearch. The system retrieves the top-k most relevant documents, ranking them based on a combination of semantic similarity, citation frequency, outcome similarity, and temporal relevance. For example, when given a criminal case involving circumstantial evidence for murder, the system can return the most cited precedents under IPC 302 and highlight the reasoning applied. This ensures that judges, lawyers, and researchers can instantly access the most legally relevant precedents without manually combing through thousands of documents.

5. Multi-Level Attention & Context Integration Module

This is the core reasoning engine of the system, responsible for integrating facts, statutes, and precedents into a unified decision space. Using advanced attention mechanisms, the module replicates the way human judges weigh evidence and law. Cross-attention layers link case facts to statutes, ensuring that factual descriptions are aligned with the legal provisions they invoke. Graph attention mechanisms reason over the law–fact graph, assigning higher importance to critical nodes such as frequently cited statutes or decisive precedents. Hierarchical attention ensures that information is aggregated across multiple levels—from individual sentences to entire judgments—while maintaining contextual relevance. By fusing multi-source information in this way, the system develops a contextually aware understanding of legal problems, enabling both accurate prediction and explainable reasoning.

6. Legal Judgment Assistance & Drafting Module

This module translates the system's reasoning into actionable legal outputs for users. It generates structured summaries of cases, including facts, issues framed, applicable statutes, arguments considered, and final

outcomes. For new queries, it can predict possible charges, relevant IPC sections, and even suggest ranges of sentences based on historical data. Beyond prediction, the module also assists in drafting judgments. Using generative language models fine-tuned on judicial writing style, it produces draft orders that follow the conventional court structure—introduction, facts, issues, reasoning, statutes applied, precedents cited, and final decision. These drafts can be used by judges or lawyers as a starting point, saving time and ensuring consistency in legal documentation.

7. User Interaction Module

A sophisticated backend is only useful if end-users can interact with it intuitively. This module provides an interface through which both legal professionals and laypersons can access the system. The interface supports natural language queries (e.g., "Find cases under IPC Section 302 involving circumstantial evidence") as well as structured input forms. The output is tailored to the user: a lawyer or judge may receive detailed precedent graphs and statute linkages, while a citizen may receive simplified explanations of charges and outcomes in plain language. Additional features include search filters by year, court, or statute, as well as visualization tools to display legal graphs and attention heatmaps. This ensures accessibility, usability, and wide adoption of the system.

8. Evaluation & Explainability Module

Given the sensitivity of legal decision-making, transparency and reliability are essential. This module is dedicated to evaluating the system's performance and ensuring explainability. The models are benchmarked using metrics such as precision, recall, F1-score (for statute prediction), and BLEU/ROUGE (for judgment text generation). Beyond quantitative metrics, the module provides interpretability through attention heatmaps that highlight which facts and statutes influenced a prediction. Users can trace how the system arrived at its outputs, building trust in its recommendations. Human-in-the-loop evaluation is also conducted, where legal experts validate the correctness of retrieved precedents, statute predictions, and generated drafts. This ensures that the system does not merely produce accurate results but also provides transparent and justifiable reasoning.

9. Cybersecurity & Data Protection Module

Legal documents often contain sensitive personal and institutional information, especially in cases related to cybercrime, financial fraud, or corporate disputes. This module ensures that data is handled with the highest standards of

cybersecurity. It incorporates encryption protocols for secure storage and transmission of case files, implements access controls for different categories of users (judges, lawyers, researchers, or citizens), and uses hashing mechanisms to maintain data integrity. Special attention is given to anonymization of personally identifiable information (PII) and sensitive digital evidence, such as IP logs, transaction records, or intercepted communications, which are commonly found in cybersecurity cases. Additionally, the system can integrate specialized modules for detecting anomalies in digital evidence, validating authenticity of electronic records under provisions like Section 65B of the Indian Evidence Act, and ensuring chain-of-custody tracking. This not only safeguards the system itself from cyber threats but also makes it capable of handling legal research in domains where cybersecurity is central.

10. High-Performance Computing & Scalability Module

Certain classes of cases, such as large-scale financial frauds, corporate insolvency disputes, or mass litigation involving thousands of documents, demand significant computational power. This module equips the system with the ability to handle such high-computation workloads efficiently. It leverages distributed computing architectures, GPU acceleration, and cloud-based elastic resources to ensure scalability. Large datasets are processed in parallel using frameworks like Apache Spark or Ray, while deep learning workloads are distributed across GPUs/TPUs for faster training and inference. The system also adopts model-parallelism and knowledge-distillation techniques to balance between performance and efficiency. By including this module, the assistant becomes suitable not only for routine case law research but also for computation-heavy legal analysis tasks such as multi-document summarization, bulk precedent retrieval, or running predictive models on thousands of concurrent cases. This ensures that the system remains reliable, responsive, and future-ready even in high-demand scenarios.

11. Evidence Management & Digital Validation Module

This module focuses on managing exhibits and evidence submitted during trials. It supports ingestion of digital evidence such as audio, video, and forensic reports, and validates them using tools like hash-checking and metadata verification. It can also cross-check whether the evidence complies with Section 65B of the Indian Evidence Act for electronic records. Judges benefit by having reliable evidence tracking, automated authenticity checks, and a structured way to refer to exhibits in judgments.

12. Courtroom Workflow Integration Module

This module is designed to integrate seamlessly with existing court management systems (like e-Courts, Case Information System, or NJDG). Judges can directly pull case files, track hearings, and push AI-assisted drafts back into their case workflow. It reduces manual steps and ensures the assistant complements existing judicial IT infrastructure.

13. Real-Time Query & Interactive Analytics Module

This allows judges to interact with the system dynamically during proceedings. For example, if a lawyer cites a precedent in court, the judge can quickly query the assistant for "other cases where Section 307 IPC was applied in similar circumstances." The module provides visual analytics (graphs, timelines, statistics) to quickly compare precedents. It acts as a real-time research companion during hearings.

14. Multi-Language Legal Understanding Module

Since many Indian cases involve regional languages, this module enables multi-lingual legal document processing. Using translation and cross-lingual embeddings, it can interpret case texts in Hindi, Tamil, Bengali, and other regional languages, aligning them with English statutes and precedents. Judges in regional courts benefit as they can access relevant precedents irrespective of the language in which they were originally written.

EXPECTED OUTCOME:

1. Improved Access to Legal Precedents

Judges will be able to instantly access the most relevant precedents without manually searching through lengthy case files. By using semantic embeddings and retrieval-augmented search, the system will identify top-cited judgments and link them to applicable statutes. This outcome reduces the time judges spend on legal research, ensures more consistent referencing of past rulings, and supports uniformity in the judicial process.

2. Enhanced Accuracy in Statute and Charge Identification

The system will automatically highlight relevant statutes and charges based on the facts of a case. For instance, if a fact pattern suggests murder under IPC 302, or cheating under IPC 420, the system will proactively retrieve and display them. This ensures that judges never overlook critical provisions and strengthens the quality of judicial reasoning by aligning it with the correct statutory framework.

3. Efficient Evidence Management and Validation

With dedicated modules for digital evidence, judges will benefit from automated checks on authenticity, compliance with Section 65B of the Evidence Act, and metadata validation. This ensures that forged, tampered, or irrelevant evidence is flagged early, reducing disputes in court. Judges gain confidence that the evidence being cited is valid and traceable, thereby reinforcing the fairness of proceedings.

4. Real-Time Decision Support in Courtrooms

The interactive analytics module will allow judges to query the system during hearings. For example, if a lawyer references a precedent, the judge can immediately verify its context, find similar cases, or review counter-arguments in real time. This outcome turns the system into a courtroom companion that augments a judge's ability to respond quickly and decisively to evolving arguments.

5. Transparent and Explainable AI Reasoning

The use of multi-level attention ensures that predictions are explainable. Judges will not only see which statutes or precedents are being suggested but also why they were chosen. Attention heatmaps and legal graphs will visually demonstrate which facts influenced the prediction, making the AI transparent. This outcome builds trust and positions the system as an aid rather than a black-box decision maker.

6. Faster Judgment Drafting and Documentation

Through the judgment drafting module, the system will provide structured drafts following the conventional legal style (facts, issues, reasoning, statutes applied, precedents cited, decision). Judges can edit or refine these drafts instead of starting from scratch. This outcome reduces paperwork burden, accelerates case disposal, and ensures judgments maintain consistency across courts.

7. Strengthened Cybersecurity in Legal Data Handling

Judges will benefit from strict data protection, encryption, and access control for case files. Sensitive personal and institutional data, especially in cybercrime and financial fraud cases, will remain protected against tampering or unauthorized access. This outcome assures that judicial data, which is highly confidential, is managed with the highest cybersecurity standards.

8. High Scalability for Complex and Bulk Cases

The system will efficiently process large-scale litigations, mass petitions, or cases involving thousands of documents by leveraging cloud-based distributed computing and GPU acceleration. Judges handling such high-computation cases will receive timely insights without delays. This outcome ensures scalability from simple single-case analysis to massive multi-party disputes.

9. Multi-Language Legal Accessibility

Judges in regional courts will gain access to precedents and judgments across multiple Indian languages. A case in Tamil can be aligned with precedents in Hindi or English through cross-lingual embeddings. This outcome promotes inclusivity, ensures equal access to justice across regions, and bridges linguistic barriers in the judiciary.

10. Reduction of Bias and Fairness Monitoring

By including fairness monitoring, the system will highlight potential biases in recommendations (e.g., skewed precedent selection). Judges can make more balanced decisions knowing that AI recommendations have been checked for neutrality. This outcome directly contributes to ethical AI adoption in the judiciary.

11. Integration with Existing Judicial Workflows

Seamless integration with e-Courts and NJDG will allow judges to access AI-assisted insights without leaving their existing workflow systems. Draft judgments, case summaries, and retrieved precedents can be pushed back into court management platforms. This outcome enhances usability and ensures smooth adoption without disrupting established practices.

12. Increased Judicial Productivity and Cost Savings

By reducing manual effort in research, evidence validation, and drafting, judges can resolve cases faster. This reduces backlog, improves disposal rates, and saves operational costs associated with prolonged litigation. The judiciary as a whole becomes more productive, delivering faster justice to citizens.

13. Improved Public Trust and Transparency

When judgments are backed by transparent AI explanations, validated evidence, and relevant precedents, citizens develop stronger trust in the judicial system. The system ensures that justice is not only done but is seen to be done through structured, explainable outputs. This outcome strengthens the legitimacy of judicial decision-making.

14. Empowerment of Judges through Decision Support

Instead of replacing judicial reasoning, the system empowers judges by acting as a research assistant. Judges remain the final decision-makers, but with enhanced knowledge support. This outcome ensures judicial independence while augmenting their ability to handle increasing caseloads with 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 proposed **AI-Driven Legal Document Research Assistant with Enhanced Context Understanding** is a transformative solution that streamlines legal research, precedent retrieval, statute identification, and judgment drafting. By integrating advanced modules such as knowledge graphs, semantic embeddings, and multi-level attention, the system bridges the gap between unstructured legal texts and structured reasoning, enabling judges to make faster, more accurate, and contextually informed decisions.

In addition to core functions, the project introduces specialized modules for **cybersecurity, high-performance computing, evidence validation, and multilingual accessibility**, making it adaptable to both routine and high-complexity cases. These features ensure that the system not only supports legal professionals in day-to-day tasks but also strengthens judicial transparency, fairness, and efficiency. At the state level, the solution contributes directly to reducing case backlogs, ensuring consistent decision-making across courts, and empowering judges with real-time decision support during proceedings.

Ultimately, the project aligns with national judicial modernization efforts such as **Digital India and e-Courts**, positioning itself as a forward-looking initiative that enhances accessibility, scalability, and trust in the legal system. Rather than replacing human wisdom, it serves as an intelligent assistant, empowering judges to focus on core reasoning while handling complex data and workflows. This ensures a justice system that is **smarter, faster, and more transparent**, delivering timely and equitable outcomes for society.

Annexure - I

Format for Appraisal of the Department

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

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	**Describe the scope of the problem that is being addressed through this proposal.** *Summarize the problem with relevant details within 2-3 paragraphs.*	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. 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. 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.
2.2	**What is the geographic, organizational, and cultural context for the problem? Who does the problem impact and how does it impact them?**	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. 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. 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.

2.3	**Please share any data, studies or articles that you have referred to, to support/evidence this problem statement.** *This could also be data from your department's work at the grassroot level or from any published journals.*	1. Anuj Kumar, Vishwas Raj, and V. S. Dixit, "NyayaAnumana: A Dataset for Legal Judgment Prediction in Indian Courts," *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1125–1137, 2023. 2 .Shubham Nigam, Aayush Bhatt, and S. Saha, "InLegalBERT: A Pre-trained Language Model for Indian Legal Texts," *International Conference on Computational Linguistics and Legal Informatics (COLIEE)*, 2022. 3. M. Chalkidis, I. Androutsopoulos, and N. Aletras, "Legal Judgment Prediction: A Survey of Approaches and Open Challenges," *Artificial Intelligence Review*, 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," *arXiv preprint*, arXiv:2307.12549, 2023. 5. N. Gupta and R. S. Yadav, "Factors Affecting Efficient Discharge of Judicial Functions in India," *Social Science Research*, vol. 113, pp. 45–58, 2024. 6. R. Bhatnagar and N. Huchhanavar, "Predicting Delays in Indian Lower Courts Using AutoML and Decision Forests," *arXiv preprint*, arXiv:2307.16285, 2023. 7. S. Mishra, "Cyclic Syndrome of Arrears and Efficiency of Indian Judiciary," *Journal of Law and Society*, 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," *Indiana University Legal Studies Research Paper*, pp. 1–28, 2011. 9. Ministry of Law and Justice, Government of India, "National Judicial Data Grid (NJDG): Statistics on Case Pendency," *Department of Justice Reports*, 2025. 10. Supreme Court of India, "Annual Report on Judicial Backlog and Case Disposal," *Supreme Court Publications*, New Delhi, 2024.
2.4	**Can this problem be addressed through the regular department budgets?**	Yes/**No**
2.5	**If no, please describe why is this problem unique and requires an innovative solution** *You may wish to describe why this cannot be considered by the regular departmental budgets and requires budgets from Special Initiative. Is the*	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. 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

	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)	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. 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.
2.6	**What is the source of the proposed solution?**	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. 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. 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.
2.7	**Are there any alternative solutions for the proposed challenge? If yes, has the department explored them?** *Please explain examples of different ways of approaching the problem that you may have explored before arriving at the proposed solution.*	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. 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.

		Another approach is **traditional digitization of court records**, which has already been explored through the **e-Courts Mission Mode Project**. While digitization enables online access to case files and judgments, it does not provide intelligent decision support. Judges still have to manually sift through hundreds of documents to identify relevant statutes and precedents, which consumes valuable time. Some pilot projects have attempted **basic search engines or keyword-based tools** for legal research. However, these systems fail to capture the semantic complexity of legal texts, resulting in either irrelevant results or overlooked precedents. They also lack explainability and the ability to validate digital evidence, making them unsuitable for high-stakes judicial decision-making. After reviewing these alternatives, it became clear that only an **AI-driven, innovation-focused solution**—combining semantic understanding, knowledge graphs, retrieval-augmented reasoning, and explainable outputs—can address the unique scale and complexity of the Indian judicial problem. Hence, the department has moved beyond traditional measures and identified this proposal as an innovative intervention worthy of special initiative support.

3	PROJECT DETAILS	
3.1	**Describe the innovation solution that is being proposed to address this challenge** *Not more than 150 words.*	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. 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**.
3.2	**Please select the type of innovation involved in your project**	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
3.3	**Please share any successful case studies of similar innovation that you may have referenced.**	**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. **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. **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 tasks. **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. **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.
3.4	**Does the proposal include a project implementation plan with a time schedule. If yes, please share it as an annexure to the proposal.**	**Yes**/No **Yes (Annexure III – Time Schedule Plan) - Enclosed**
3.5	**If no, please share the tentative date by when this will be submitted**	---
4	**PROJECT OUTCOMES**	
4.1	**Describe the economic benefits proposed through the implementation of this project** *Not more than 100 words*	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.
4.2	**Is the project scalable through your department on successful implementation of the pilot?**	Yes

4.3	**If yes, indicate the department's plans**	• 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 after the TANII funding period?**	Yes/No Yes, the project includes continued financial sustainability plans after the TANII funding period.
4.5	**If yes, indicate the departments plans**	The Department Plans are: • 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.**	**₹ 2,75,00,000 (Rupees Two Crore Seventy-Five Lakh Only)**
5.2	**Please indicate the project duration** *(Minimum of 1 year and Maximum of 3 yrs allowed)*	**3 Years**
5.3	**Other sources of funding if available**	Institutional support - using the existing facilities for the development of the project
5.4	**Cost of the machinery included in the project**	**High-Performance Servers / GPU Workstations** – For training and deploying AI/ML models (₹ 40–50 Lakhs) **Cloud Infrastructure Setup** – Hybrid cloud/on-premise storage, distributed compute clusters (₹ 25–30 Lakhs) **Secure Storage Systems** – Encrypted storage servers for legal documents and case files (₹ 10–12 Lakhs) **Networking & Security Appliances** – Firewalls, VPNs, and intrusion detection systems to ensure cybersecurity (₹ 5–7 Lakhs) **Backup & Disaster Recovery Systems** – High-capacity storage devices and redundancy hardware (₹ 8–10 Lakhs) **User Interface Hardware** – Courtroom terminals, secured tablets, or kiosks for judge/staff access (₹ 5–8 Lakhs)
5.5	**Percentage of the Machinery cost to the project cost**	33 %
5.6	**Cost of the Computers and Accessories included in the project**	90 Lakhs
5.7	**Percentage of the Computers and Accessories included in the project**	33 %

5.8	**Is the purchase of new machinery indispensable for Implementation of the project?**	Yes/No Yes 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. 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.
5.9	**Cost of experts / consultants included in the project cost**	15 lakhs
5.1	**Percentage of the Cost of experts / consultants to the project cost**	5.45 %
5.11	**Is the budgetary support requested under the proposed innovation within the prescribed ceiling of Rs.10 crores and proposed annual expenditure within the limit of Rs.5 crores?**	Yes/No Yes, the budgetary support requested for the proposed innovation project is well within the prescribed ceiling of ₹10 crores, with the **total project cost at ₹2.75 crores** and the **annual expenditure comfortably below ₹5 crores**.
5.12	**Does the proposal include break up of budgets as Capital Expenditure and Revenue Expenditure?** *Please note that all TANII projects are required to submit this break up as the funds will be sanction through separate head of accounts.*	Yes/No Yes, 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. Capital Expenditure (CapEx): ₹1,20,35,100 Revenue Expenditure (RevEx): ₹1,54,64,900 Total: ₹2,75,00,000

5.13	**If no, please share the tentative date by when this will be submitted**	----
6	**Others**	
6.1	**Does the project require Statutory clearances**	Yes / No *(if 'Yes', the details should be furnished)* **No**
6.2	**Statutory clearances obtained**	Yes / No *(if 'No', the reasons to be furnished)* **No**
6.3	**Assessment on possible risks and challenges**	Yes / No ● **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 language and graph models need advanced servers, GPUs, and scalable infrastructure. ● **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?**	Yes / No Yes, the project aligns with several national and institutional initiatives that focus on technology-driven legal research and justice delivery. ● 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**	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.
6.6	**Resources required to complete the project preparation process** *(Please detail: need to engage consultants including transaction advisors)*	To complete the project preparation process, the resources required are: • **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 computing (HPC) clusters, secure cloud storage, and GPU resources, will be required to train and deploy large legal language models. • **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**	Roles and Responsibilities of Involved Parties
		• Project Implementing Agency (Lead Institution / University)
		  ○ 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)
		  ○ 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) ○ 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 ○ 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) ○ Provision high-performance computing, GPU clusters, and secure cloud storage. ○ Ensure uptime, scalability, and cybersecurity of deployed platforms. ● Stakeholders (Judges, Lawyers, Law Students, Researchers) ○ 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	85,00,000		85,00,000
Sector Specific/Area of focus related machinery or equipment	50,00,000		50,00,000
Computers, Electronics and other accessories	40,00,000		40,00,000
Revenue Expenditure			
Software and other digital assets	25,00,000		25,00,000
Manpower	55,00,000		55,00,000
Travel	10,00,000		10,00,000
Other cost (specify)	10,00,000		10,00,000
Total	**2,75,00,000**	**0**	**2,75,00,000**

**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	40,00,000			30,00,000			15,00,000		85,00,000
Sector Specific/Area of focus related machinery or equipment	25,00,000			15,00,000			10,00,000		50,00,000
Computers, Electronics and other accessories	20,00,000			15,00,000			5,00,000		40,00,000
Revenue Expenditure									
Software and other digital assets	10,00,000			8,00,000			7,00,000		25,00,000
Manpower	20,00,000			20,00,000			15,00,000		55,00,000
Travel	4,00,000			3,00,000			3,00,000		10,00,000
Other cost (specify) – Training, Consumables, and Maintenance	5,00,000			3,00,000			2,00,000		10,00,000
Total	**1,24,00,000**	**0**	**0**	**94,00,000**	**0**	**0**	**57,00,000**	**0**	**2,75,00,000**

**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*
**Please note that all TANII projects have an upper cost limit of INR 500 lakhs per year and INR 1000 lakhs overall*

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	BIZON G9000 / equivalent – 4× NVIDIA A100/H100/H200 Tensor Core AI GPU Server		42,50,000	2	85,00,000
2	**Sector Specific/Area of focus related machinery or equipment**	High-Speed Document Scanners with OCR (200 ppm) Legal Data Processing Servers (for NLP + graph construction)		7,50,000 17,50,000	2 2	15,00,000 35,00,000
3	**Computers, Electronics and other accessories**	Developer Workstations (Intel Xeon/AMD Threadripper, 128GB RAM, 2TB SSD) Audio/Video Equipment for Annotation & Voice Processing Networking Devices (NAS, Routers, Edge Servers)		5,00,000 2,50,000 10,00,000	5 2 1 set	25,00,000 5,00,000 10,00,000
Total						1,75,00,000

**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 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**	AI/ML Software Licence Legal NLP / Case-law Database Access Cloud Storage & Compute Credits (AWS / NIC Cloud)		2,50,000 1,50,000 3,00,000	7,00,000 (INR / year)	21,00,000
2	**Manpower**	Project Scientist Project Associate 1 Project Assistant		60000 38000 26000	16,08,000 (INR / year)	48,24,000
3	**Travel**	Court Visits / Data Collection Training & Awareness Workshops National Conferences / Seminars		2,00,000 1,00,000 50,000	3,50,000 (INR / year)	10,50,000
4	**Other cost (specify) Training, Consumables, and Maintenance**	Workshops Data Acquisition Connectivity Equipment maintenance Consumables		1,60,000 1,20,000 1,60,000 1,40,000	5,80,000 (INR / year)	17,40,000
TOTAL						1,00,00,000

**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*
**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

	ACTIVITIES	Year 1				Year 2				Year 3			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
FOUNDATION & PROTOTYPING													
1	**Stakeholder Engagement & Dataset Preparation**												
	Design an AI-legal research architecture, and Plan dataset acquisition & preprocessing (OCR, tagging, normalization)												
	System Architecture & Prototyping												
	Develop a high-performance legal AI server system, Prototype web/app interface for legal research, and Prototype a voic assistant for multilingual legal queries.												
	Collect baseline legal data: case texts, statutes, sentencing patterns; Capture early document scans (PDF/OCR); Start gathering voice queries in regional languages.												
AI MODEL DEVELOPMENT, FIELD TESTING & CONTEXT ANALYTICS													
2.	Develop AI models for: Statute prediction, Charge identification and Sentence estimation; Case relevance detection usin embeddings/graph reasoning; Voice analytics system: Speech-to text for legal queries, Intent recognition, and Text-to-speec outputs.												
	Integration & Pilot Deployment												
	Integrate all modules (document ingestion, AI models, graph reasoning, voice assistant); Deploy the pilot version in selected courts/law schools; Train stakeholders to use the web/voice interface; Connect the AI engine to legal search and retrieval systems.												
	Continuous Data Collection												
	Collect case texts, voice queries, and document data; Record user queries and AI responses for NLP tuning; Gather system performance metrics (accuracy, precision, recall, adoption).												
OPTIMIZATION, SCALING & EVALUATION													
3.	**Model Tuning & System Optimization**												
	Refine AI models with Year 2 case data; Fine-tune NLP models using voice query logs; Optimize research interface usability an system performance.												
	Scaling & Expansion												
	Expand deployment to additional courts and legal institutions; Customize AI recommendations for different jurisdictions and case types; Establish collaborations with Bar Councils and law schools for adoption.												
	Impact Assessment												
	Evaluate system performance: Improvement in research efficiency, Accuracy of statute/charge/sentence predictions, Use satisfaction and adoption rate; Perform cost–benefit analysis.												
	Documentation & Dissemination												
	Conduct workshops and prepare training materials & technical manuals; Publish research articles, showcase demos, and prepar final project report.												