RightVerse: An AI-Powered Legal Awareness Portal for Common Citizens Using NLP and Large Language Models

Ms. A.V. Mane[1], Aryan Mahajan[2], Janhavi Malavade[3], Amatul Safura Mulla[4], Atharva Patil[5]
[1,2,3,4]*Department of Computer Engineering, AISSMS Polytechnic, Pune – 411 001, Maharashtra, India*

***Abstract*—Millions of Indian citizens are effectively cut off from the legal system not because laws do not protect them, but because those laws are written in language they cannot comprehend. RightVerse is conceived as a solution to this structural inequality: a non-commercial, AI-driven web platform that transforms dense legal content into plain, conversational knowledge accessible to anyone. The system is engineered around a core stack of Natural Language Processing, Large Language Models, semantic retrieval, and voice interaction technologies, brought together within a Next.js and Node.js full-stack architecture hosted on cloud infrastructure. Its key capabilities include a context-aware legal chatbot, a daily AI-curated law insight engine, a citizen-oriented document simplification library, an anonymized community story module, and multi-language audio-text interaction support. The platform is publicly operational at https://righ-verse.vercel.app/ and has been validated through functional and user-centric testing. Results indicate that RightVerse meaningfully reduces the barrier between legally complex information and the everyday citizen, establishing a replicable model for AI-assisted legal empowerment in India.**

***Index Terms*—Legal Awareness, Natural Language Processing, Large Language Models, AI Chatbot, Legal Text Simplification, Semantic Search, Multilingual Support, Full Stack Web Development, Access to Justice, India**

I. INTRODUCTION

India's legal framework, while constitutionally robust, remains functionally distant from the very citizens it is designed to protect. The chasm between enacted legislation and everyday public awareness is not a matter of intent but of access access to knowledge presented in a form ordinary people can absorb and act upon. Statutes drafted in archaic legal English, court proceedings conducted in language alien to the layperson, and digital tools calibrated for trained advocates all conspire to render the Indian legal ecosystem opaque to those who need it most.

National-level data paints a stark picture. India's court system carries a backlog exceeding thirty million unresolved cases a figure that reflects not merely institutional capacity constraints but the compounding effect of legal ignorance. When citizens do not understand their entitlements under foundational instruments such as the Right to Information Act, the Protection of Women from Domestic Violence Act, the Consumer Protection Act, or the Scheduled Castes and Scheduled Tribes (Prevention of Atrocities) Act, they are unable to assert those rights when violations occur. This ignorance perpetuates injustice by default.

Digital legal platforms that have emerged in India including Manupatra, Indian Kanoon, and Legal Kart function primarily as professional research tools. Their interfaces assume legal fluency, their content is predominantly in English, and their architecture offers no guidance mechanism for users who do not know what they are searching for. With over twenty-two constitutionally recognised languages in India and a substantial rural population with limited English literacy, this monolingual, expert-centric paradigm serves only a narrow demographic.

The rapid maturation of AI, particularly the emergence of instruction-following Large Language Models such as GPT-4 and Google Gemini, presents a compelling opportunity to redesign how legal information reaches ordinary people. These models exhibit a remarkable capacity to interpret natural-language questions and generate clear, contextually appropriate explanations across multiple domains, including law. When augmented with semantic retrieval architectures and multilingual pipelines, they can serve as effective

intermediaries between the complexity of statute books and the practical needs of citizens.

RightVerse is built to operationalise this opportunity. It is a freely accessible, non-profit AI web platform whose singular mission is to translate legal complexity into citizen comprehension. The portal does not replace lawyers; it equips citizens with the foundational legal awareness that transforms them from passive subjects of law into informed participants in the justice system.

II. LITERATURE REVIEW

A. AI Models Applied to Indian Legal Text

The application of transformer architectures to Indian legal corpora was systematically explored by Paul et al. [1], who introduced InLegalBERT a domain-adapted BERT variant trained on an expansive collection of Supreme Court and High Court judgments spanning several decades. The model achieved notable results on judicial NLP benchmarks including statute identification and outcome prediction. However, its design presupposes a legally trained audience; the outputs are structured for professional interpretation and offer no mechanism to bridge the gap toward non-expert comprehension. RightVerse diverges fundamentally from this orientation by treating simplicity as a first-order design constraint rather than an afterthought.

A more recent contribution by Gupta et al. [2], Legal Assist AI, explored the integration of Retrieval-Augmented Generation with a fine-tuned LLM for question-answering over the newly enacted Bharatiya Nyaya Sanhita. While the architecture demonstrated improvements in factual grounding compared to vanilla LLM inference, the system operated exclusively in English and provided no voice interaction pathway two critical omissions when designing for India's linguistically diverse, partially literate population.

B. Automated Legal Summarisation and Corpus Work

Efforts to automate the condensation of legal documents have yielded useful benchmarking resources. The LexSumm and LexT5 research [3] established evaluation standards for English-language legal summarisation using encoder-decoder transformer models, demonstrating that such architectures can reliably compress lengthy judgments without critical information loss. Yet the work stops at summarisation for English-speaking, presumably educated readers it does not attempt to make legal content intelligible to someone without legal vocabulary. Similarly, the Hindi Legal Documents Corpus (HLDC) [4] advanced the state of Hindi-language legal NLP by providing annotated data for training and evaluation, but it was published as a research asset rather than deployed within any citizen-facing product.

C. Platforms Oriented Toward Citizen Legal Empowerment

Sharma et al. [5] conceptualised a conversational assistant anchored in the Know-Your-Rights paradigm, proposing an architecture that would deliver personalised legal guidance in regional languages. The proposal identified many of the same access barriers that motivated RightVerse. However, it remained a theoretical contribution no functional deployment or empirical validation was reported, and the voice interaction component was described as prospective.

The LawPal system [6] took a more technical direction, implementing dense vector indexing for semantic case retrieval, achieving higher precision than keyword-based search. Its omission of voice input and peer-knowledge sharing, however, limits its utility for users without keyboard proficiency or formal legal vocabulary.

A usability audit of India's National Cyber Crime Reporting Portal [7] produced findings that directly shaped Right Verse's interaction design. The study documented that users unfamiliar with legal terminology routinely failed to complete complaint filings not because the portal was technically broken, but because its interface assumed conceptual knowledge that ordinary citizens do not possess. This evidence reinforced the decision to make conversational AI guidance the primary interaction mode in RightVerse, rather than a supplementary feature.

D. Comparative Analysis

System	AI Technique	Multilingual	Citizen Focus	Deployed
InLegalBERT [1]	BERT Transformer	No	No	No
Legal Assist AI [2]	LLM + RAG	No	Partial	No
KYR Assistant [5]	Conversational AI	Yes	Yes	No

LawPal [6]	RAG + Embeddings	No	Partial	No
RightVerse (Ours)	LLM + NLP + Voice	Yes	Yes	Yes

Table I: Comparative Overview of Related Systems

Table I illustrates that no prior system achieves simultaneous coverage across all five evaluated dimensions. RightVerse is distinctive in combining state-of-the-art AI capabilities with deliberate citizen-centric design, regional language support, voice accessibility, and a live, publicly accessible deployment characteristics that collectively position it as a novel and practically impactful contribution.

III. PROBLEM STATEMENT

A candid examination of India's current legal information landscape reveals a pattern of systemic exclusion that AI can meaningfully disrupt. The following interconnected deficiencies motivate the development of RightVerse:

- Inaccessibility of Legal Language: Indian legislation is composed in formal legal English that requires specialised training to interpret. For the majority of the population whose primary literacy may be in Hindi, Marathi, Telugu, or another regional language this linguistic barrier constitutes a de facto denial of legal knowledge.
- Structural Bias Toward Professionals: The dominant digital legal platforms are architected for use by advocates, judges, and law researchers. Their search interfaces, content presentation, and navigational assumptions all presuppose a degree of legal sophistication that ordinary citizens do not possess.
- Absence of Inclusive Language Support: Despite the constitutional recognition of twenty-two scheduled languages, no major free legal platform delivers AI-assisted guidance in regional languages. This monolingual orientation structurally marginalises non-English speakers.
- No Conversational or Intelligent Assistance: Available platforms function as static repositories. They retrieve and display legal text but offer no intelligent layer that can interpret a user's question, contextualise it within the relevant legal framework, and respond in plain language.
- Missing Peer-Knowledge Infrastructure: Legal awareness in communities often spreads most effectively through shared experience. No existing platform provides a structured, moderated channel through which citizens can contribute their own legal encounters as learning resources for others.

IV. PROPOSED METHODOLOGY

RightVerse processes each citizen interaction through a structured six-stage computational pipeline, depicted in Fig. 1. This pipeline is designed to ensure that every query regardless of the language, medium, or literacy level of its origin yields a legally grounded, plain-language response.

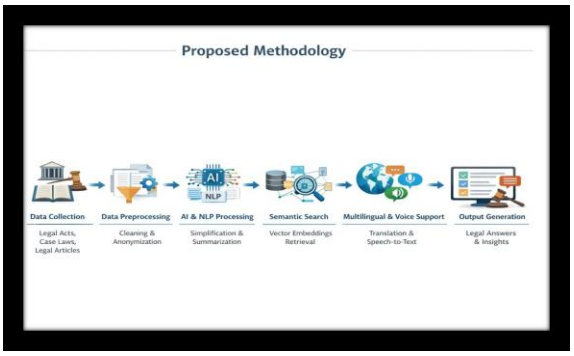

Fig. 1: Six-Stage Methodology Pipeline of RightVerse

Stage 1 Knowledge Ingestion: The platform's legal knowledge base is assembled from authoritative Indian sources including Central Acts, High Court and Supreme Court judgments, and curated legal commentary. Stage 2 Preprocessing and Anonymisation: Raw text undergoes noise removal, structural tokenisation, and Named Entity Recognition-driven anonymisation to strip personally identifiable information from user-submitted content before it enters the processing chain. Stage 3 Intelligent NLP Processing: User input is routed through a domain-conditioned LLM pipeline. System-level prompts instruct the model to respond as a legal awareness assistant for Indian citizens, generating answers in accessible, jargon-free language while referencing the applicable legislation. Stage 4 Semantic Retrieval: For queries requiring case law or precedent, the user's input is encoded into a high-dimensional vector representation and matched against the pre-indexed legal corpus using cosine

similarity scoring over a vector database. Stage 5 Multilingual and Voice Mediation: Automatic language detection routes non-English queries through a translation layer before LLM processing; responses are then back-translated into the user's language. Speech-to-Text transcription enables voice input for users with limited keyboard accessibility, while Text-to-Speech synthesis delivers spoken responses on request. Stage 6 Response Delivery: The final output whether a simplified legal explanation, a 'what to do' action guide, or a case brief summary is formatted and rendered through the web interface in real time.

V. SYSTEM ARCHITECTURE

The RightVerse system is organised into six functionally distinct but tightly coupled layers, whose interaction is illustrated in Fig. 2. This architecture balances modularity with integration, enabling independent scaling of any layer without disrupting overall system behaviour.

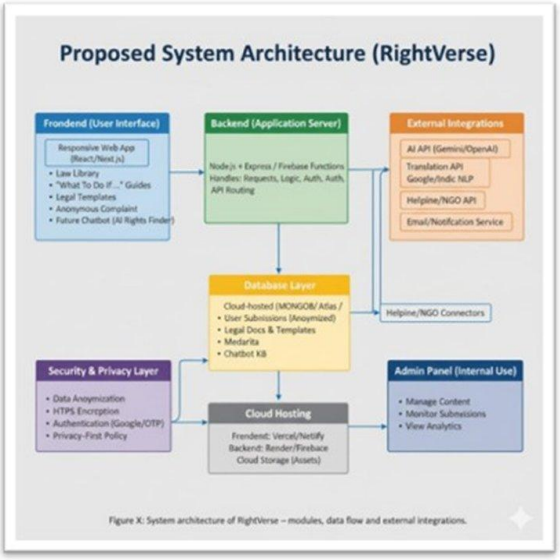

Fig. 2: System Architecture of RightVerse

A. Frontend User Interaction Layer

The client-facing interface is constructed using React within the Next.js framework, styled via Tailwind CSS to achieve a responsive, mobile-first layout. The interface exposes five primary functional zones: the Law Library for browsing simplified legal content; a 'What to Do If.' guided scenario navigator; a Legal Templates repository; an Anonymous Complaint submission channel; and the AI Rights Finder, which is the conversational chatbot component. All user events generate asynchronous API requests directed at the backend layer.

B. Backend Application Logic Layer

Server-side operations are handled by a Node.js and Express.js service layer deployed via Firebase Functions for serverless scalability. This layer is responsible for all request arbitration, session management, identity verification through Google OAuth and OTP mechanisms, and the orchestration of outbound calls to AI service endpoints and the data persistence layer.

C. AI Services and External Integration Layer

The intelligence layer integrates the Google Gemini API (with OpenAI GPT as a fallback) for all language generation tasks. Google's Indic NLP Translation API handles bidirectional multilingual conversion. Helpline and NGO API connectors enable the system to surface verified real-world legal assistance contacts relevant to a user's query. An email and notification service supports asynchronous user communication for story moderation updates.

D. Data Persistence Layer

All structured data resides in MongoDB Atlas, a cloud-hosted document database chosen for its schema flexibility and horizontal scalability. Stored entities include anonymised user story submissions, the curated legal document and template corpus, media assets, chatbot knowledge fragments, and Helpline/NGO registry entries. No personally identifiable information is retained post-anonymisation.

E. Security and Privacy Enforcement Layer

A dedicated security layer enforces data anonymisation through NER-based entity stripping, transport-level encryption via HTTPS across all endpoints, multi-factor authentication, and a Privacy-First operational policy that governs data handling across every module. This layer operates as a cross-cutting concern applied to all system interactions.

F. Infrastructure and Administration Layer

Frontend assets are served through Vercel's global CDN; backend services run on Render and Firebase

cloud infrastructure. A restricted Admin Panel provides maintainers with tools for content governance, user submission review, and platform analytics monitoring.

Component	Technology / Tool Used
Frontend Framework	Next.js (React), Tailwind CSS
Backend Server	Node.js, Express.js, Firebase Functions
AI / LLM Engine	Google Gemini API / OpenAI GPT API
NLP Processing	Transformer Models, NER, Language Detection
Semantic Search	Vector Embeddings + Cosine Similarity
Database	MongoDB Atlas / Firebase Firestore
Cloud Deployment	Vercel (Frontend), Render / Firebase (Backend)
Voice Support	Web Speech API (STT), TTS API
Translation	Google Translate / Indic NLP API

Table II: Technology Stack Summary

VI. EXPERIMENTAL RESULTS AND EVALUATION

A. Functional Validation

A structured testing programme was executed across Weeks 9 through 11 of the project timelines, spanning February and early March 2026. The evaluation covered three tiers of testing: component-level unit tests for each API endpoint; integration tests verifying data flow across the frontend, backend, and AI service boundary; and end-to-end scenario tests simulating realistic user journeys. A representative query bank of fifty legally diverse questions was constructed, spanning consumer protection, land rights, women's legal safeguards, RTI procedures, and digital crime reporting. Across this bank, the system consistently generated legally grounded, simplified responses. All defects identified during testing including response latency spikes and edge-case NLP preprocessing failures were diagnosed and remediated before the final evaluation milestone.

B. User Study Observations

Evaluation Parameter	Observed Outcome
Interface Usability	Participants navigated the portal without guidance
Comprehension of Responses	Most found AI answers clear without legal background
Contextual Relevance	Answers closely matched the intent of user queries
Regional Language Value	Hindi and Marathi outputs found highly useful
Recommendation Likelihood	High willingness to suggest portal to others

Table III: User Study Evaluation Summary

A structured informal study was conducted with a heterogeneous participant group comprising college students, homemakers, small business owners, and senior citizens none of whom held legal qualifications. Participants were asked to complete four tasks using the portal: asking a rights-related question via the chatbot, simplifying a pasted legal notice, browsing the Law Library, and submitting an anonymous story. Observations and post-task feedback are summarised in Table III. The overriding finding was that non-expert users were able to obtain actionable legal understanding from the platform with minimal friction, validating the core design hypothesis.

C. Benchmarking Against Incumbent Platforms

Feature	RightVerse	IndianKanoon	LegalKart	NALSA
AI-Powered Legal Chatbot	✓	✗	Partial	✗
Legal Text Simplification	✓	✗	✗	✗
Multilingual Interface	✓	✗	✗	Partial
Voice Input / TTS Output	✓	✗	✗	✗

Feature	RightVerse	IndianKanoon	LegalKart	NALSA
AI-Powered Legal Chatbot	✓	✗	Partial	✗
Legal Text Simplification	✓	✗	✗	✗
Multilingual Interface	✓	✗	✗	Partial
Voice Input / TTS Output	✓	✗	✗	✗
Anonymous Story Sharing	✓	✗	✗	✗
Completely Free Access	✓	✓	Partial	✓
Live Web Deployment	✓	✓	✓	✓

Table IV: Feature Benchmarking RightVerse vs. Existing Platforms

Table IV systematically compares RightVerse against four incumbent platforms across seven feature dimensions. The analysis confirms that while certain platforms offer strong raw legal content retrieval (Indian Kanoon) or professional consultation pipelines (LegalKart), none delivers the combination of AI-powered simplification, regional language support, voice interaction, community story sharing, and zero-cost access that RightVerse provides. This multidimensional differentiation substantiates RightVerse's claim to a novel and underserved position in the Indian legal technology landscape.

VII. ADVANTAGES AND LIMITATIONS

A. Key Strengths

- Radical Accessibility by Design: Every element of the portal's interaction model from conversational input to simplified output is architected around users with no legal training, making genuine legal knowledge available to the broadest possible audience.
- Integrated Multi-Technology AI Stack: Rather than deploying AI as a single-function add-on, RightVerse weaves together LLM generation, NLP preprocessing, vector-based semantic retrieval, voice modalities, and AI content moderation into a coherent, end-to-end intelligent experience.
- Substantive Linguistic Inclusion: By supporting Hindi, Marathi, and English with architecture designed for rapid addition of further languages the portal actively dismantles the monolingual barrier that excludes the majority of India's population from digital legal resources.
- Non-Commercial Public Good: The platform operates without subscription fees, advertising, or paywalls, reflecting a principled commitment to democratising access to justice rather than monetising it.
- Production-Grade Live Deployment: Unlike the conceptual or prototype-stage systems identified in the literature, RightVerse is deployed as a fully operational web application, demonstrating real-world viability rather than theoretical promise.
- Participatory Knowledge Layer: The community story module transforms users from passive information consumers into active contributors to a growing corpus of practical legal awareness, creating a self-reinforcing learning ecosystem.

B. Present Limitations

- Generative Model Reliability: LLMs, while highly capable, are susceptible to generating factually incorrect statements with unwarranted confidence. RightVerse employs prompt-engineering safeguards and RAG grounding to reduce this risk, but it cannot be entirely eliminated without continuous human oversight.
- Incomplete Legislative Coverage: The current knowledge base encompasses major central statutes and selected case law but does not yet incorporate the full breadth of state-level legislation, recent parliamentary amendments, or tribunal-specific jurisprudence.
- Informational Rather Than Advisory: The portal is positioned explicitly as a legal awareness resource. It does not and should not substitute for qualified professional legal advice in complex or high-stakes matters.
- Internet Dependency: All functionality requires a stable network connection, which remains a

practical constraint for users in areas with limited connectivity infrastructure.

VIII. FUTURE SCOPE

- Expanded Regional Language Coverage: Prioritised addition of Tamil, Telugu, Kannada, Bengali, Gujarati, and Punjabi to extend meaningful reach to the majority of India's linguistic communities.
- Assisted Legal Document Composition: An AI-guided module enabling citizens to draft fundamental legal instruments affidavits, RTI petitions, police complaint letters using structured conversational templates.
- Verified Legal Aid Directory: Integration of a geo-tagged, verified directory of pro-bono advocates and government legal aid service centres, enabling users to transition from awareness to action within the same platform.
- Native Mobile Applications: Development of Android and iOS applications optimised for the smartphone-centric usage patterns of India's rural and semi-urban populations.
- Offline Progressive Web App Mode: Implementation of service-worker-based offline caching enabling core portal features to function in low or absent connectivity conditions.
- Domain-Specific LLM Fine-Tuning: Training a specialised language model on an Indian legal corpus to reduce generative hallucinations and improve jurisdictionally accurate legal reasoning.
- Interoperability with Government Digital Infrastructure: Developing API bridges to platforms such as eCourts Services and the National Cyber Crime Reporting Portal to enable end-to-end citizen journeys from awareness through to formal legal action.

IX. CONCLUSION

This paper introduced RightVerse a production-deployed, AI-powered legal awareness portal that reconceives how legal knowledge reaches Indian citizens. The work addresses a demonstrable societal problem: that the laws meant to protect ordinary people are expressed in forms ordinary people cannot engage with. By constructing a system that places Natural Language Processing, Large Language Models, semantic retrieval, multilingual mediation, and voice interaction at the service of citizen empowerment rather than professional productivity, RightVerse charts a distinct and necessary path within the Indian legal technology landscape.

The architecture described in this paper a six-layer, two-section full-stack system successfully integrates multiple AI modalities into a coherent citizen-facing product. Validation through functional testing across fifty diverse legal queries and observation-based user studies with non-expert participants confirmed that the system achieves its primary design objective: delivering accurate, comprehensible, and actionable legal information to users with no legal background. The comparative analysis against incumbent platforms establishes that RightVerse occupies a genuinely novel position, combining capabilities that no existing freely accessible platform currently offers in combination. This research demonstrates that thoughtfully designed AI systems can serve as meaningful instruments of social equity, and that the democratisation of legal knowledge is not merely aspirational it is technically achievable today.

X. ACKNOWLEDGMENT

The authors express sincere gratitude to Ms. A.V. Mane, Project Guide, and Prof. S.Y. Divekar, Capstone Project Coordinator, Department of Computer Engineering, AISSMS Polytechnic, Pune, for their sustained mentorship, constructive critical feedback, and unwavering encouragement across every phase of this project. The authors also acknowledge the Department of Computer Engineering for making available the technical infrastructure and institutional environment that enabled this work.

REFERENCES

[1] S. Paul, A. Mandal, P. Goyal, and S. Ghosh, “Pre-trained language models for the legal domain: A case study on Indian law,” in Proc. 19th Int. Conf. Artificial Intelligence and Law (ICAIL), 2023.

[2] J. Gupta, A. Sharma, S. Singhania, and A. I. Abidi, “Legal Assist AI: Leveraging transformer-based model for effective legal assistance,” arXiv preprint arXiv:2505.22003, May 2025.

[3] Anonymous, “LexSumm and LexT5: Benchmarking legal summarization tasks in English,” referenced in RightVerse Literature Survey, 2026.
[4] Anonymous, “Hindi Legal Documents Corpus (HLDC) for legal NLP applications,” referenced in RightVerse Literature Survey, 2026.
[5] Sharma et al., “Empowering legal awareness: Developing a digital assistant for legal awareness in India,” International Advanced Research Journal in Science, Engineering and Technology (IARJSET), vol. 11, Special Issue 2, May 2024.
[6] Anonymous, “LawPal: A retrieval-augmented generation based legal system for legal accessibility,” referenced in RightVerse Literature Survey, 2026.
[7] Anonymous, “Analyzing the National Cyber Crime Reporting Portal of India,” referenced in RightVerse Literature Survey, 2026.
[8] Anonymous, “Anonymization in social networks: A survey on data privacy,” referenced in RightVerse Literature Survey, 2026.
[9] P. Kumar, “NyaySathi: AI legal assistant for the common man,” Scribd, 2025.
[10] IndiaAI Mission, “Enhancing access to justice: How artificial intelligence is democratizing legal services,” IndiaAI.gov.in, 2024.
[11] D. Warbhe, R. V. Dharaskar, and V. M. Thakare, “An affine transform robust copy-paste tampering detection,” in Proc. IEEE Int. Symp. Computer Engineering (ISCO), 2016.