

JUST-NLP 2025

The 1st Workshop on NLP for Empowering Justice

Proceedings of the Workshop

December 24, 2025

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-312-8

Introduction

We are pleased to share the Proceedings of JUST-NLP 2025: The First Workshop on NLP for Empowering Justice, which was held concurrently with IJCNLP-AAACL 2025. This workshop represents a significant step forward in strengthening the connection between Natural Language Processing (NLP) and the legal field. This is especially true in places like India, where numerous languages are spoken, numerous cases need to be resolved, and access to legal services is limited. Researchers, practitioners, and technologists who work at the cutting edge of Legal-NLP were supposed to meet at JUST-NLP. Legal NLP is an area that requires methodological rigor, domain sensitivity, and careful consideration of real-world, high-stakes applications.

Obtaining legal information remains a challenge worldwide. In the Indian judiciary, particularly in lengthy case decisions, the use of complex legal reasoning structures and the fact that English is the primary language used in official proceedings make it challenging to ensure transparency and fairness. Automating key aspects of legal information processing, such as summarization, translation, retrieval, and reasoning, could make tasks easier for lawyers and citizens. The primary goal of JUST-NLP 2025 was to accelerate progress in these areas and establish a robust research community centered on Legal-NLP.

Overall, our first edition featured a research track, four invited talks, and two shared tasks: (1) L-SUMM, an abstractive summarization task for Indian legal judgments, and (2) L-MT, a legal machine translation task between English and Hindi. The introduction of two shared tasks, each addressing important needs in legal information processing, was a highlight of this year’s workshop. The workshop received 29 submissions, of which 21 were accepted, highlighting a strong community interest in Legal NLP and its applications to the legal processing pipeline. Among the accepted papers, 5 were regular research-track papers published in the proceedings, and 2 were accepted as non-archival presentations. For the shared tasks, 9 papers were accepted for L-SUMM, and 5 papers were accepted for L-MT, for publication in the proceedings. Since the event was held in a hybrid format, presentations were delivered both in person at the IJCNLP-AAACL 2025 venue in Mumbai and virtually.

The Shared Task on Summarization of Indian Court Judgments (L-SUMM) addressed the challenge of creating concise, clear, and accurate summaries of lengthy legal texts. It added the InLSum dataset, which has 1,800 court decisions and expert-written, abstract summaries of those decisions. Nine teams that participated examined a wide range of methods, including long-context transformers, hierarchical and multi-stage summarization frameworks, extractive-abstractive hybrids, rhetorical-role-aware chunking, and reinforcement-learning-based alignment. The results demonstrate the importance of long-context modeling and adapting to specific fields in creating effective legal summaries.

The Shared Task on English-to-Hindi Legal Machine Translation (L-MT) also aimed to make India’s multilingual legal system more accessible. Using a carefully chosen 50,000-sentence English–Hindi legal parallel corpus and a thorough evaluation framework, the participating teams tested multilingual Transformers, QLoRA-based parameter-efficient fine-tuning, curriculum learning, reinforcement learning with verifiable rewards, and even Transformer models trained from scratch. The results show that domain-adaptive finetuning and precision-focused optimization are effective for getting accurate legal translations. This shows that MT systems can be useful in real-world legal workflows.

In addition to the shared tasks, the workshop included research contributions on understanding legal texts, reasoning with statutes and case law, extracting information from specific domains, mining arguments, adapting legal-domain LLMs, and using AI for justice in a socially responsible way. The papers in this volume demonstrate that this research area is maturing, but they also highlight some problems that still need to be addressed, such as ensuring long-form generation is factually consistent, reducing hallucinations in high-stakes fields, and incorporating expert-driven evaluation protocols.

Our workshop also featured a diverse set of invited talks from leading experts across law, industry, public institutions, and AI research. The talks highlighted multiple perspectives on the future of Legal NLP, ranging from legal theory and courtroom-scale AI deployment to ontology-driven knowledge representations and technology-enabled governance reform.

We believe that the resources, ideas, and discussions emerging from JUST-NLP 2025 will set the stage for future progress at the intersection of law and language technologies. As NLP systems improve, they are more likely to make legal knowledge accessible to everyone, making it easier for all to access justice. We hope that this workshop will foster greater collaboration, responsible innovation, and a deeper understanding of the unique challenges that Legal-NLP presents.

We would like to thank the authors, shared-task organizers, program committee members, reviewers, and participants for their hard work and assistance. We would also like to thank the individuals who helped organize the IJCNLP–AAACL 2025 event. We are excited to see JUST-NLP continue to grow in the years to come.

Organizing Committee

Program Chairs

Ashutosh Modi, IIT Kanpur, India
Saptarshi Ghosh, IIT Kharagpur, India
Asif Ekbal, IIT Patna, India
Pawan Goyal, IIT Kharagpur, India
Sarika Jain, NIT Kurukshetra, India

Web and Publicity Chairs

Abhinav Joshi, IIT Kanpur, India
Shivani Mishra, IIT Kanpur, India

Shared Task and Publication Chairs

Debtanu Datta, IIT Kharagpur, India
Shounak Paul, IIT Kharagpur, India
Kshetrimayum Boynao Singh, IIT Patna, India
Sandeep Kumar, IIT Patna, India

Program Committee

Program Committee

Abhinav Joshi, IIT Kanpur, India
Abhisek Sharma, NIT Kurukshetra, India
Amita Jain, NSUT Delhi, India
Aniket Deroy, IIT Delhi, India
Debtanu Datta, IIT Kharagpur, India
Kshetrimayum Boynao Singh, IIT Patna, India
Pooja Harde, NIT Kurukshetra, India
Sandeep Kumar, IIT Patna, India
Shivani Mishra, IIT Kanpur, India
Shounak Paul, IIT Kharagpur, India
Sumit Dalal, Bennett University, India

Invited Speakers

Niraj Kumar, NLU Delhi, India
Arghya Bhattacharya, Adalat AI, India
Joseph Pookkatt, Staram, India
Sampritha Manjunath, Insight SFI Centre for Data Analytics, Ireland
Parth Parikh, eSuccess AI Technologies, India
Nishi Yadav, Ministry of Tribal Affairs, Government of India, India

Keynote Talk

Artificiality of Law in reference to Artificial Intelligence

Niraj Kumar
NLU Delhi, India

Abstract: This talk examines the evolving relationship between legal theory and emerging AI systems, drawing on insights from constitutional and comparative law. It explores how AI challenges traditional conceptions of legal reasoning, authority, and legitimacy, and raises foundational questions about the artificiality of law itself. The talk also reflects on the implications of entrusting interpretive and decision-support roles to artificial agents within legal systems.

Bio: Prof. Niraj Kumar is a Professor of Law and Director of the Centre for Comparative Law at National Law University Delhi, India. His areas of expertise include Constitutional Law, Administrative Law, Comparative Law, Environmental Law, and Legal Theory. He previously served as Additional Registrar (Research) at the Supreme Court of India, attached with Hon'ble the Chief Justice of India from January 2019 to May 2021. He has authored several notable books, including *The Indian Legal System: An Enquiry* (Oxford, 2019) and multiple volumes of the *Indian Yearbook of Comparative Law* (Springer). His research has also been published in leading journals, including the *Journal of the Indian Law Institute*, *NUJS Law Review*, and *Shimla Law Review*. He has also been a resource person for training officers from the IAS, IPS, Judicial Services, and other civil services.

Keynote Talk

Building AI for India's Courtrooms: Scaling Sovereign Voice and Language Systems Across 20 percent of the Judiciary

Arghya Bhattacharya
Adalat AI, India

Abstract: India is home to the world's largest and most linguistically complex justice system – one that processes millions of hearings every day across dozens of languages and procedural contexts. Over the last two years, Adalat AI has deployed voice and language AI in nearly 20% of all Indian courtrooms, powering live transcription, dictation, translation, and document navigation at a population scale. This talk walks through the journey of building these systems end-to-end: how we approached the machine-learning stack for courts, how we handle legal context across diverse tasks, and what it takes to engineer sovereign, privacy-preserving AI for public institutions. I'll share the technical and operational challenges we encountered, speech variability, multilingual legal phrasing, courtroom acoustics, edge-case reasoning, and deployment constraints, and the architectural decisions that allowed us to scale from pilots to thousands of courtrooms. The session will highlight lessons on building domain-grounded AI for high-stakes environments, discuss why voice AI is only one part of a much broader AI for Justicestack, and outline what it means to build reliable, inclusive, and future-proof digital public goods for the Global South.

Bio: Mr. Arghya Bhattacharya is the Co-Founder and CTO of Adalat AI, where he leads the development of sovereign voice and language technologies that now power courtrooms across India. A researcher and engineer by training, he holds a Bachelor's in Computer Science and a Master's in Artificial Intelligence from IIIT Hyderabad, with publications in leading venues including ACL, EMNLP, CoNLL, and EAMT. Before launching Adalat AI, Arghya was the first founding engineer at Enterpret, where he built large-scale NLP systems for customer intelligence, and later worked at Equal on identity-verification infrastructure for high-security environments. His academic work in multilingual and low-resource NLP underpins Adalat AI's vernacular court-workflow stack. In 2024, he co-founded Adalat AI with Utkarsh Saxena to modernize India's justice system through AI. Today, their platform supports live transcription, dictation, translation, and legal analytics across thousands of courts. In 2025, Arghya was named to the Forbes 30 Under 30 Asia – Social Impact list for advancing accessible, rights-aligned AI for public institutions. His broader mission is to build sovereign, privacy-preserving digital public goods that enhance inclusion, strengthen state capacity, and support justice systems across the Global South.

Keynote Talk

Designing and Implementing Knowledge Graphs in the Legal Domain

Joseph Pookkatt, Sampritha Manjunath, Parth Parikh

Staram, India; Insight SFI Centre for Data Analytics, Ireland; eSuccess AI Technologies, India

Abstract: Ontologies play a crucial role in enabling data integration from multiple sources by providing a common vocabulary for diverse systems. This is particularly important for AI applications, which often rely on data from various databases and formats. Ontologies align this disparate data, making seamless processing possible. They also empower AI to perform semantic searches, interpreting user intent rather than just matching keywords. This allows users to make more natural language queries, enabling AI to retrieve precise and meaningful information. In the realm of Retrieval-Augmented Generation (RAG), the dominant approach has been BaselineRAG. However, BaselineRAG has a notable limitation: it struggles to connect the dots between concepts that are semantically distant but logically connected. This is where Knowledge Graphs (KGs) excel. By exploring relationships within the graph using graph embeddings or graph-specific algorithms, GraphRAG captures the context and logical relationships between concepts. This makes it far better positioned to deliver higher-quality, more accurate outputs compared to traditional RAG methods. In the context of Indian law, building a large, unified knowledge graph has historically been too challenging due to the heterogeneous systems involved. Staram is tackling this issue by customizing multiple knowledge graphs for specific legal use cases in India. They will demonstrate how Large Language Models (LLMs) can infer graph architectures and symbolic layers tailored to specific needs. This approach enables the creation of smaller, specialized graphs that are accurate, interoperable, and easier to govern. These advancements hold significant promise for building generative AI applications that transform legal research, compliance, and decision-making.

Bio: 1. Joseph Pookkatt: Mr. Joseph Pookkatt is a practicing lawyer with over 30 years of experience at the intersection of corporate law, litigation, legal technology, AI, and entrepreneurship. As a co-founder of Staram Analytics, Joseph collaborates with legal policy organizations to develop ontologies for specific laws and is leading the development of India's first open-source legal knowledge graph. His work aims to make Indian legal systems more accessible, efficient, and transparent through the power of AI and ontologies.

2. Sampritha Manjunath: Ms. Sampritha Manjunath is an NLP Data Scientist with 9+ years of experience in language models, knowledge graphs, RAG, and conversational AI. Experienced in leading research projects, developing scalable AI solutions, and publishing in top venues such as ACL and LDK. Currently a Research Associate at the University of Galway, focused on building impactful and responsible AI systems.

3. Parth Parikh: Mr. Parth Parikh is an AI Research Engineer at eSuccess AI Technologies, where he leads work in legal NLP, knowledge graphs, and ontology-driven AI for large-scale legal reasoning. He earned his M.Tech in Computer Science from IIT Bombay (2023), with research focused on applied machine learning and Computer Vision. At eSuccess AI, Parth works on building advanced legal intelligence systems powered by LLMs and retrieval-augmented architectures. His current interests include structured legal representations, legal data models, and domain-adapted language models.

Keynote Talk

Fixing the Process, Powering the System: What Jharkhand Teaches India About Legal-Tech

Nishi Yadav

Ministry of Tribal Affairs, Government of India, India

Abstract: Access to justice is a core democratic promise, yet its delivery in India remains deeply strained. With 4.7 crore cases pending across courts, delays routinely turn justice delayed into justice denied. The Government, Centre and States combined, is the single largest litigant, responsible for nearly half of all cases, including 46% of High Court matters, many of which arise from service-related disputes in departments like Education, Railways, Health, and Finance. This volume of litigation imposes severe costs: rising legal expenditure, reduced administrative bandwidth, and declining institutional trust. In several departments, administrators spend over 40% of their time managing court matters. Right from 1974 by Justice Krishna Iyer, the 100th Law Commission Report - 1983, the 126th Law Commission Report – 1988, to the Prime Minister leading meetings between the Department of Justice with the Central Government, and the more recent attempt to accelerate a National Litigation Policy, the issue has been acknowledged and attempts have been made at reform. A consistent theme in all of these efforts has been that of wanting to reduce litigation involving the Govt. Experience across states shows that government litigation is rarely only a legal problem. It stems from incoherent policies, weak grievance redressal, and limited institutional capacity for managing legal processes. Inconsistent decisions, poor tracking of judicial precedents, and the absence of mechanisms to resolve disputes internally often push routine administrative issues into courts. This systemic challenge sets the stage for why Jharkhand's legal-tech transformation is particularly notable. The Case of Jharkhand: While digital governance conversations often highlight Telangana, Karnataka, or Kerala, one of India's most quietly impactful reforms has unfolded in Jharkhand - a young, resource-rich, predominantly tribal state. Its legal-tech journey stands out for prioritising process redesign before technology adoption. Jharkhand followed a pragmatic approach: map processes, fix bottlenecks, then digitise. This ensured that technology supported well-defined workflows instead of automating inefficiencies. The state's journey began with the Vidhi Portal (2016) in the Advocate General's Office, which evolved into the central platform for real-time service of notices and orders. Building on this foundation, the Integrated Litigation Management System (ILMS) launched in 2020 by the Department of School Education & Literacy made the department genuinely data-driven. Case pendency dropped sharply; response times shrank from 45–70 days to just 7 days; and administrative culture shifted from reactive firefighting to proactive governance. This cooperative ecosystem, spanning the AG Office, the Department of IT & e-Governance, and key line departments, stands in contrast to national systems like LIMBS, which function largely as data-entry repositories with limited integration or decision-support capability. Jharkhand is now preparing its next leap: AI-enabled predictive governance to identify recurring disputes, strengthen policy design, and reduce litigation at its root. This talk positions Jharkhand's experience not just as a state-level success but as a scalable blueprint for India, where governance reform begins with process, matures with technology, and evolves with intelligence.

Bio: An Economics graduate from DU and a lawyer with over 11 years of PQE, Ms. Nishi began her career engaging with underserved communities through organizations like WGHR-UN, NLUD, and HRLN. Her tenure at the Chief Minister's Office, Delhi Government, ignited her passion for understanding and improving government systems. Since 2018, she has been working closely with the Government of Jharkhand and various non-profit organisations, leading transformative initiatives to drive policy reforms and strengthen governance, developing expertise in analyzing legal processes from constitutional, technological, and organizational perspectives to improve efficiency and effectiveness. She spearheaded the study,

design, development, and implementation of the Integrated Litigation Management System (ILMS) with the Government of Jharkhand which automated file movement processes, reducing the pendency of court cases from 11,000 to 3,500. Alongside, she improved key processes such as leave management, teacher transfers, and school monitoring, integrating legal-tech solutions with system-wide reforms. Additionally, she has conducted extensive studies in close association with bureaucracy in other states, combining legal acumen, technological insights, and strategic vision. Committed to socio-economic and governance reforms, her collaborations with various state governments highlight her ability to navigate the intricacies of public systems and officers while building systems that drive meaningful impact. She is currently involved in strategic governance and legal-tech projects with the Ministry of Tribal Affairs (GoI), the Government of Jharkhand in collaboration with IIT-Kanpur and IIT (ISM) Dhanbad, and ongoing reform initiatives with the Government of Jammu & Kashmir.

Table of Contents

Overview of the 1st Workshop on NLP for Empowering Justice

Ashutosh Modi, Saptarshi Ghosh, Asif Ekbal, Pawan Goyal, Sarika Jain, Abhinav Joshi, Shivani Mishra, Debtanu Datta, Shounak Paul, Kshetrimayum Boynao Singh and Sandeep Kumar 1

Findings of the JUST-NLP 2025 Shared Task on Summarization of Indian Court Judgments

Debtanu Datta, Shounak Paul, Kshetrimayum Boynao Singh, Sandeep Kumar, Abhinav Joshi, Shivani Mishra, Sarika Jain, Asif Ekbal, Pawan Goyal, Ashutosh Modi and Saptarshi Ghosh 5

Findings of the JUST-NLP 2025 Shared Task on English-to-Hindi Legal Machine Translation

Kshetrimayum Boynao Singh, Sandeep Kumar, Debtanu Datta, Abhinav Joshi, Shivani Mishra, Shounak Paul, Pawan Goyal, Sarika Jain, Saptarshi Ghosh, Ashutosh Modi and Asif Ekbal 12

LeCNet: A Legal Citation Network Benchmark Dataset

Pooja Harde, Bhavya Jain and Sarika Jain 18

Legal Document Summarization: A Zero-shot Modular Agentic Workflow Approach

Taha Sadikot and Sarika Jain 29

LLM Driven Legal Text Analytics: A Case Study For Food Safety Violation Cases

Suyog Joshi, Soumyajit Basu, Lipika Dey and Partha Pratim Das 41

Grahak-Nyay: Consumer Grievance Redressal through Large Language Models

Shrey Ganatra, Swapnil Bhattacharyya, Harshvivek Kashid, Spandan Anaokar, Shruthi N Nair, Reshma Sekhar, Siddharth Manohar, Rahul Hemrajani and Pushpak Bhattacharyya 53

Nyay-Darpan: Enhancing Decision Making Through Summarization and Case Retrieval for Consumer Law in India

Swapnil Bhattacharyya, Harshvivek Kashid, Shrey Ganatra, Spandan Anaokar, Reshma Sekhar, Shruthi N Nair, Siddharth Manohar, Rahul Hemrajani and Pushpak Bhattacharyya 73

Cold Starts and Hard Cases: A Two-Stage SFT-RLVR Approach for Legal Machine Translation (Just-NLP L-MT shared task)

Pawitsapak Akarajardwong and Chompakorn Chaksangchaichot 101

Contextors at L-SUMM: Retriever-Driven Multi-Generator Summarization

Pavithra Neelamegam and S Jaya Nirmala 107

A Budget Recipe for Finetuning a Long-form Legal Summarization Model

Chompakorn Chaksangchaichot and Pawitsapak Akarajardwong 113

SCaLAR_NITK @ JUSTNLP Legal Summarization (L-SUMM) Shared Task

Arjun T D and Anand Kumar Madasamy 121

goodmen @ L-MT Shared Task: A Comparative Study of Neural Models for English-Hindi Legal Machine Translation

Deeraj S K, Karthik Suryanarayanan, Yash Ingle and Pruthwik Mishra 127

NIT-Surat@L-Sum: A Semantic Retrieval-Based Framework for Summarizing Indian Judicial Documents

Nita Jadav, Ashok Urlana and Pruthwik Mishra 133

Adapting IndicTrans2 for Legal Domain MT via QLoRA Fine-Tuning at JUST-NLP 2025

Akoijam Jenil Singh, Loitongbam Sanayai Meetei and Yumnam Surajkanta 142

<i>Team-SVNIT at JUST-NLP 2025: Domain-Adaptive Fine-Tuning of Multilingual Models for English–Hindi Legal Machine Translation</i>	
Rupesh Dhakad, Naveen Kumar and Shrikant Malviya	148
<i>Combining Extractive and Generative Methods for Legal Summarization: BLANCKED at JUST-NLP 2025</i>	
Erich Giuseppe Soto Parada, Carlos Manuel Muñoz Almeida and David Cuevas Alba	155
<i>Automatic Legal Judgment Summarization Using Large Language Models: A Case Study for the JUST-NLP 2025 Shared Task</i>	
Santiago Chica	162
<i>Structure-Aware Chunking for Abstractive Summarization of Long Legal Documents</i>	
Himadri Sonowal and Saisab Sadhu	171
<i>From Scratch to Fine-Tuned: A Comparative Study of Transformer Training Strategies for Legal Machine Translation</i>	
Amit Barman, Atanu Mandal and Sudip Kumar Naskar	179
<i>Integrating Graph based Algorithm and Transformer Models for Abstractive Summarization</i>	
Sayed Ayaan Ahmed Sha, Sangeetha Sivanesan, Anand Kumar Madasamy and Navya Binu . .	186
<i>Hierarchical Long-Document Summarization using LED for Legal Judgments</i>	
Reshma Sheik, Noah John Puthayathu, Fathima Firose A and Jonathan Paul	191