NLLP 2025

Natural Legal Language Processing Workshop 2025

Proceedings of the Workshop

November 8, 2025

The NLLP organizers gratefully acknowledge the support from the following sponsors.

Gold

Bloomberg

Engineering

HUMAN ADS

erc

European Research Council

©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 979-8-89176-338-8

Introduction

Welcome to the seventh edition of the NLLP (Natural Legal Language Processing) Workshop, co-located with the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Different industrial sectors have embraced natural language processing (NLP) technologies, which have altered services and products in healthcare, finance, education among others. The legal domain, in particular, is at the frontier for new tools, tasks and exploration for new interesting research problems. Electronic tools have become increasingly integral in the practice of law, with their usage projected to undergo exponential growth. By its very nature, the practice of law necessarily involves the analysis and interpretation of language. The potential for NLP applications to provide benefit to practitioners of law and consumers of legal services around the world is enormous and is already manifesting itself.

We organized this workshop to bring together researchers, practitioners, policy makers from around the world who develop NLP techniques within the legal domain. This is an exciting opportunity to expand the boundaries of our field by identifying new problems and exploring new data as it interacts with the full inventory of NLP and machine learning approaches. In this spirit, the Organizing and Program Committee was assembled to include researchers from both academia and industry, from NLP and legal backgrounds.

We were interested in the following types of papers: (1) applications of NLP to legal tasks; (2) methods for applying Large Language Models (LLMs) to the legal domain; (3) methodological innovations for legal tasks; (4) tasks, resources and evaluation for NLP in the Legal domain; (5) NLP for online platforms, social media and regulation; (6) systems, demos and industry applications with legal applications and (7) interdisciplinary position papers on law and NLP.

We once again received an exceptionally high number of submissions. In total, we received 71 submissions on our direct submission website, out of which we accepted 29 papers for an acceptance rate of 40%. We also received 7 ARR (ACL Rolling Review) commitments -papers that have received reviews and meta-review via the ARR system- of which we accepted 3 for publication. Overall, we accepted 32 papers out of 78 submissions (overall acceptance rate is 41%), out of which 4 papers were submitted as non-archival. All papers were reviewed by at least 3 members of the Program Committee, one of whom was usually a legal scholar.

The papers in this year's edition of the workshop cover a wide range of topics including new data sets for legal NLP in lower resource languages such as Portuguese or Thai, benchmarks of LLMs on specific challenges in legal documents such as long context and applications of LLMs and NLP insights for tasks on legal documents. Applications include retrieval augmented generation, Personal Identifiable Information (PII) identification, summarization, argumentation, semantic parsing, long-context inference, information retrieval, classification, prompt optimization and multi-agent systems. Again, we are thrilled to see the variety of legal documents that are the subject of papers in the workshop including contracts, wills, patents, court cases, statues, terms of services, trademarks and copyright infringement documents. We note a specific focus this year on reasoning models and their application to legal documents.

We thank everyone who expressed interest in the workshop, all authors of submitted papers, members of the Program Committee who did an excellent job at reviewing papers given a short turnaround time, everyone attending the workshop and the EMNLP 2025 conference for hosting the workshop. We thank our sponsors: Bloomberg and the European Research Council Starting Grant project HUMANads.

The NLLP Workshop organizers.

http://nllpw.org

Organizing Committee

Organizers

Nikolaos Aletras, The University of Sheffield
Ilias Chalkidis, University of Copenhagen
Leslie Barrett, Bloomberg Law
Cătălina Goanță, Utrecht University
Daniel Preoțiuc-Pietro, Bloomberg
Gerasimos Spanakis, Maastricht University

Program Committee

Reviewers

Sallam Abualhaisa, Tomaso Agnoloni, Ion Androutsopoulos, Tom Ault

Ilayda Balaban, Claire Barale, Thales Bertaglia, Floris Bex, Baldwin Breck

Luca Cagliero, Jiahong Chen, Odysseas Spyridon Chlapanis, Ashish Chouhan

Bram Duivenvoorde, Dominik Dworniczak, Arthur Dyevre

Dimitrios Galanis, Piyush Ghai

Ivan Habernal, Ben Hagag, Nils Holzenberger, Abe Hou

Abderrahmane Issam

Baaij Jaap

Constantinos Karouzos, Aykut Koc, Alice Kwak

Tong Liang, Ruta Liepina, Chu Luo

Megan Ma, Pawel Maka, Adam Meyers, Jelena Mitrović

Rohan Nanda, Joel Niklaus

Henrik Palmer Olsen, Katsikouli Panagiota, Ioannis Panagis, Anu Pradhan

Paulo Quaresma

T.Y.S.S. Santosh, Vageesh Saxena, Yusuf Can Semerci, Gil Semo, Madhavan Seshadri, Samyak Sheth, Dan Simonson, Jerrold Soh, Christoph Sorge, Alexandru Sotropa, Ieva Staliunaite

Dimitrios Tsarapatsanis, Kalpana Tyagi

Gijs Van Dijck, Helen van der Kroef

Jianqiang Wang, Hannes Westermann

ShanShan Xu, Huiyin Xue

Marcos Zampieri, Miri Zilka, Frederike Zufall

Table of Contents

Tracing Definitions: Lessons from Alliance Contracts in the Biopharmaceutical Industry
Maximilian Kreutner, Doerte Leusmann, Florian Lemmerich and Carolin Haeussler 1

The Automated but Risky Game: Modeling Agent-to-Agent Negotiations and Transactions in Consumer Markets
Shenzhe Zhu, Jiao Sun, Yi Nian, Tobin South, Alex Pentland and Jiaxin Pei 16

Towards Reliable Retrieval in RAG Systems for Large Legal Datasets
Markus Reuter, Tobias Lingenberg, Ruta Liepina, Francesca Lagioia, Marco Lippi, Giovanni Sartor, Andrea Passerini and Burcu Sayin 17

Translating Tax Law to Code with LLMs: A Benchmark and Evaluation Framework
Gabriele Lorenzo, Aldo Pietromatera and Nils Holzenberger 31

Beyond the Haystack: Sensitivity to Context in Legal Reference Recall
Eric Xia, Karthik Srikumar, Keshav Karthik, Advaith Renjith and Ashwinee Panda 48

Machine Unlearning of Personally Identifiable Information in Large Language Models
Dan Parii, Thomas van Osch and Chang Sun 54

Modeling Motivated Reasoning in Law: Evaluating Strategic Role Conditioning in LLM Summarization
Eunjung Cho, Alexander Miserlis Hoyle and Yoan Hermstrüwer 68

Label-Free Distinctiveness: Building a Continuous Trademark Scale via Synthetic Anchors
Huihui Xu and Kevin D. Ashley 113

Copyright Infringement by Large Language Models in the EU: Misalignment, Safeguards, and the Path Forward
Noah Scharrenberg and Chang Sun 125

Grounded Answers from Multi-Passage Regulations: Learning-to-Rank for Regulatory RAG
Tuba Gokhan and Ted Briscoe 135

NyayGraph: A Knowledge Graph Enhanced Approach for Legal Statute Identification in Indian Law using Large Language Models
Siddharth Shukla, Tanuj Tyagi, Abhay Singh Bisht, Ashish Sharma and Basant Agarwal 147

Evaluating LLM-Generated Legal Explanations for Regulatory Compliance in Social Media Influencer Marketing
Haoyang Gui, Thales Bertaglia, Taylor Annabell, Catalina Goanta, Tjomme Dooper and Gerasimos Spanakis 157

Validate Your Authority: Benchmarking LLMs on Multi-Label Precedent Treatment Classification
M. Mikail Demir and M Abdullah Canbaz 172

Labor Lex: A New Portuguese Corpus and Pipeline for Information Extraction in Brazilian Legal Texts
Pedro Vitor Quinta de Castro and Nádia Félix Felipe Da Silva 184

Evaluating the Role of Verifiers in Test-Time Scaling for Legal Reasoning Tasks
Davide Romano, Jonathan Richard Schwarz and Daniele Giofrè 207

Domain Adapted Text Summarization with Self-Generated Guidelines
Andrianos Michail, Bartosz Rudnikowicz, Pavlos Fragkogiannis and Cristina Kadar 226

PILOT-Bench: A Benchmark for Legal Reasoning in the Patent Domain with IRAC-Aligned Classification Tasks
Yehoon Jang, Chaewon Lee, Hyun-seok Min and Sungchul Choi 240

Efficient Prompt Optimisation for Legal Text Classification with Proxy Prompt Evaluator
Hyunji Lee, Kevin Chenhao Li, Matthias Grabmair and Shanshan Xu 281

ContractEval: Benchmarking LLMs for Clause-Level Legal Risk Identification in Commercial Contracts
Shuang Liu, Zelong Li, Ruoyun Ma, Haiyan Zhao and Mengnan Du 291

Contemporary LLMs struggle with extracting formal legal arguments
Lena Held and Ivan Habernal 292

Aligning LLMs for Thai Legal Question Answering with Efficient Semantic-Similarity Rewards
Pawitsapak Akarajaradwong, Chompakorn Chaksangchaichot, Pirat Pothavorn, Ekapol Chuangsuwanich, Attapol Rutherford and Sarana Nutanong 304

Not ready for the bench: LLM legal interpretation is unstable and uncalibrated to human judgments
Abhishek Purushothama, Junghyun Min, Brandon Waldon and Nathan Schneider 317

LeMAJ (Legal LLM-as-a-Judge): Bridging Legal Reasoning and LLM Evaluation
Joseph Enguehard, Morgane Van Ermengem, Kate Atkinson, Sujeong Cha, Arijit Ghosh Chowdhury, Prashanth Kallur Ramaswamy, Jeremy Roghair, Hannah R Marlowe, Carina Suzana Negreanu, Kitty Boxall and Diana Mincu 318

A Framework to Retrieve Relevant Laws for Will Execution
Md Asiful Islam, Alice Saebom Kwak, Derek Bambauer, Clayton T Morrison and Mihai Surdeanu 338

CourtNav: Voice-Guided, Anchor-Accurate Navigation of Long Legal Documents in Courtrooms
Sai Khadloya, Kush Juvekar, Arghya Bhattacharya and Utkarsh Saxena 351

Are LLMs Court-Ready? Evaluating Frontier Models on Indian Legal Reasoning
Kush Juvekar, Arghya Bhattacharya, Sai Khadloya and Utkarsh Saxena 359

LegalSim: Multi-Agent Simulation of Legal Systems for Discovering Procedural Exploits
Sanket Badhe 370

Linking Transparency and Accountability: Analysing The Connection Between TikTok's Terms of Service and Moderation Decisions
Leonard Eßer and Gerasimos Spanakis 382

Risks and Limits of Automatic Consolidation of Statutes
Max Prior, Adrian Hof, Niklas Wais and Matthias Grabmair 396

GReX: A Graph Neural Network-Based Rerank-then-Expand Method for Detecting Conflicts Among Legal Articles in Korean Criminal Law
Seonho An, Young-Yik Rhim and Min-Soo Kim 408

GuRE:Generative Query REwriter for Legal Passage Retrieval
Daehui Kim, Deokhyung Kang, Jonghwi Kim, Sangwon Ryu and Gary Lee 424

Extract-Explain-Abstract: A Rhetorical Role-Driven Domain-Specific Summarisation Framework for Indian Legal Documents
Veer Chheda, Aaditya Uday Ghaisas, Avantika Sankhe and Dr. Narendra Shekokar 439

Program

Saturday, November 8, 2025

09:00 - 17:00 *Papers*

Tracing Definitions: Lessons from Alliance Contracts in the Biopharmaceutical Industry
Maximilian Kreutner, Doerte Leusmann, Florian Lemmerich and Carolin Haeussler

The Automated but Risky Game: Modeling Agent-to-Agent Negotiations and Transactions in Consumer Markets
Shenzhe Zhu, Jiao Sun, Yi Nian, Tobin South, Alex Pentland and Jiaxin Pei

Towards Reliable Retrieval in RAG Systems for Large Legal Datasets
Markus Reuter, Tobias Lingenberg, Ruta Liepina, Francesca Lagioia, Marco Lippi, Giovanni Sartor, Andrea Passerini and Burcu Sayin

Translating Tax Law to Code with LLMs: A Benchmark and Evaluation Framework
Gabriele Lorenzo, Aldo Pietromatera and Nils Holzenberger

Beyond the Haystack: Sensitivity to Context in Legal Reference Recall
Eric Xia, Karthik Srikumar, Keshav Karthik, Advaith Renjith and Ashwinee Panda

Machine Unlearning of Personally Identifiable Information in Large Language Models
Dan Parii, Thomas van Osch and Chang Sun

Modeling Motivated Reasoning in Law: Evaluating Strategic Role Conditioning in LLM Summarization
Eunjung Cho, Alexander Miserlis Hoyle and Yoan Hermstrüwer

Label-Free Distinctiveness: Building a Continuous Trademark Scale via Synthetic Anchors
Huihui Xu and Kevin D. Ashley

Copyright Infringement by Large Language Models in the EU: Misalignment, Safeguards, and the Path Forward
Noah Scharrenberg and Chang Sun

Grounded Answers from Multi-Passage Regulations: Learning-to-Rank for Regulatory RAG
Tuba Gokhan and Ted Briscoe

NyayGraph: A Knowledge Graph Enhanced Approach for Legal Statute Identification in Indian Law using Large Language Models
Siddharth Shukla, Tanuj Tyagi, Abhay Singh Bisht, Ashish Sharma and Basant Agarwal

Evaluating LLM-Generated Legal Explanations for Regulatory Compliance in Social Media Influencer Marketing
Haoyang Gui, Thales Bertaglia, Taylor Annabell, Catalina Goanta, Tjomme Dooper and Gerasimos Spanakis

Validate Your Authority: Benchmarking LLMs on Multi-Label Precedent Treatment Classification
M. Mikail Demir and M Abdullah Canbaz

Labor Lex: A New Portuguese Corpus and Pipeline for Information Extraction in Brazilian Legal Texts
Pedro Vitor Quinta de Castro and Nádia Félix Felipe Da Silva

Evaluating the Role of Verifiers in Test-Time Scaling for Legal Reasoning Tasks
Davide Romano, Jonathan Richard Schwarz and Daniele Giofrè

Domain Adapted Text Summarization with Self-Generated Guidelines
Andrianos Michail, Bartosz Rudnikowicz, Pavlos Fragkogiannis and Cristina Kadar

PILOT-Bench: A Benchmark for Legal Reasoning in the Patent Domain with IRAC-Aligned Classification Tasks
Yehoon Jang, Chaewon Lee, Hyun-seok Min and Sungchul Choi

Efficient Prompt Optimisation for Legal Text Classification with Proxy Prompt Evaluator
Hyunji Lee, Kevin Chenhao Li, Matthias Grabmair and Shanshan Xu

ContractEval: Benchmarking LLMs for Clause-Level Legal Risk Identification in Commercial Contracts
Shuang Liu, Zelong Li, Ruoyun Ma, Haiyan Zhao and Mengnan Du

Contemporary LLMs struggle with extracting formal legal arguments
Lena Held and Ivan Habernal

Aligning LLMs for Thai Legal Question Answering with Efficient Semantic-Similarity Rewards
Pawitsapak Akarajaradwong, Chompakorn Chaksangchaichot, Pirat Pothavorn, Ekapol Chuangsuwanich, Attapol Rutherford and Sarana Nutanong

Not ready for the bench: LLM legal interpretation is unstable and uncalibrated to human judgments
Abhishek Purushothama, Junghyun Min, Brandon Waldon and Nathan Schneider

LeMAJ (Legal LLM-as-a-Judge): Bridging Legal Reasoning and LLM Evaluation
Joseph Enguehard, Morgane Van Ermengem, Kate Atkinson, Sujeong Cha, Arijit Ghosh Chowdhury, Prashanth Kallur Ramaswamy, Jeremy Roghair, Hannah R Marlowe, Carina Suzana Negreanu, Kitty Boxall and Diana Mincu

A Framework to Retrieve Relevant Laws for Will Execution
Md Asiful Islam, Alice Saebom Kwak, Derek Bambauer, Clayton T Morrison and Mihai Surdeanu

CourtNav: Voice-Guided, Anchor-Accurate Navigation of Long Legal Documents in Courtrooms
Sai Khadloya, Kush Juvekar, Arghya Bhattacharya and Utkarsh Saxena

Are LLMs Court-Ready? Evaluating Frontier Models on Indian Legal Reasoning
Kush Juvekar, Arghya Bhattacharya, Sai Khadloya and Utkarsh Saxena

LegalSim: Multi-Agent Simulation of Legal Systems for Discovering Procedural Exploits
Sanket Badhe

Linking Transparency and Accountability: Analysing The Connection Between TikTok's Terms of Service and Moderation Decisions
Leonard Eßer and Gerasimos Spanakis

Risks and Limits of Automatic Consolidation of Statutes
Max Prior, Adrian Hof, Niklas Wais and Matthias Grabmair

GReX: A Graph Neural Network-Based Rerank-then-Expand Method for Detecting Conflicts Among Legal Articles in Korean Criminal Law
Seonho An, Young-Yik Rhim and Min-Soo Kim

GuRE:Generative Query REwriter for Legal Passage Retrieval
Daehui Kim, Deokhyung Kang, Jonghwi Kim, Sangwon Ryu and Gary Lee

Extract-Explain-Abstract: A Rhetorical Role-Driven Domain-Specific Summarisation Framework for Indian Legal Documents
Veer Chheda, Aaditya Uday Ghaisas, Avantika Sankhe and Dr. Narendra Shekokar