

NLLP 2024

Natural Legal Language Processing Workshop 2024

Proceedings of the Workshop

November 16, 2024

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

Gold

Bloomberg

Engineering

HUMAN
ADS



©2024 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 979-8-89176-183-4

Introduction

Welcome to the sixth edition of the NLLP (Natural Legal Language Processing) Workshop, co-located with the 2024 Conference on Empirical Methods in Natural Language Processing.

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, stands as a promising frontier for the exploration of 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.

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 methods to legal tasks; (2) experimental results using and adapting NLP methods in legal documents; (3) descriptions of new legal tasks for NLP; (4) creation of curated and/or annotated resources; (5) descriptions of systems which use NLP technologies for legal text; (6) industrial research in this area and (7) interdisciplinary position papers.

In addition, this year we introduced The LegalLens Shared Task, organized by Darrow.ai in collaboration with the NLLP2024 workshop including two shared sub-tasks: Sub-Task A. Legal Named Entity Recognition (L-NER): Given possible online media text (review), determine or extract legal entities such as violation, "violation by," violation on, and law. Sub-Task B. Legal Natural Language Inference (L-NLI): Given a premise summarizing a class action complaint and a hypothesis from an online media text, determine if the relationship is entailed, contradicted, or neutral, indicating any association between the review and the complaint. A total of 87 individual users grouped in 38 teams participated in the shared task, out of which the seven highest-scoring teams were elected to write a system description paper. Most of the teams participated in both sub-tasks.

We once again received an exceptionally high number of submissions. In total, we received 68 submissions on our direct submission website, out of which we accepted 26 for an acceptance rate of 38%. We also received 5 ARR (ACL Rolling Review) commitments -papers that have received reviews and meta-review via the ARR system- of which we accepted 2 for publication. Overall, we accepted 28 papers out of 73 submissions (overall acceptance rate remains 38%) All papers were reviewed by at least 3 members of the Program Committee, one of whom was usually a legal scholar. In addition, we also offered a venue for presentation for 5 papers accepted to the Findings of EMNLP 2024 on the above topics. Finally, the proceedings feature 7 system description papers for the shared task as well as one paper that describes the shared task, for a total of 41 papers. All papers except shared task descriptions were invited to have an oral presentation, either in-person or remote and all papers were invited to participate in the poster session.

The papers cover a wide range of topics including new data sets for legal NLP, position papers on open legal NLP issues, legal perspectives on NLP topics, as well as applications of NLP tasks to legal documents including retrieval, information extraction, generation, legal judgement prediction and statutory reasoning. Again we saw a number of papers explore the use of Large Language Models (LLMs) in the legal domain spanning from fundamental questions to usage methods and applications such as prompting strategies, chain-of-thought applications in retrieval tasks or e-discovery. We note in particular a focus on language model evaluation this year and new legal corpora that we hope will lead to a synergy promoting better modeling and better testing on legal texts going forward.

We thank our invited speaker Omri Ben-Shahar, who is the Leo and Eileen Herzel Professor of Law, and

Kearney Director and founder of the Coase-Sandor Institute for Law and Economics at the University of Chicago Law School for accepting our invitation. In the tradition of past NLLP workshops, the invited speaker is a legal scholar with an interest in empirical methods for legal analysis including NLP methods. His talk is entitled Privacy Protection, At What Cost?which aims to challenge the adequacy of the data privacy protection paradigm and expose the harms to society that certain privacy-fueled protections may cause.

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 2024 conference for hosting us and the workshop. We thank our sponsors – Bloomberg and the European Research Council Starting Grant project HUMANads – for their contributions.

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, Elliott Ash, Tom Ault

Claire Barale, Thales Bertaglia, Andrew Blair-Stanek, Stephanie Brandl, Baldwin Breck

Rajarithnam Chandramouli, Jiahong Chen, Odysseas Spyridon Chlapanis, Ashish Chouhan

Marina Danilevsky, Tony Davis, Stefania Degaetano-Ortlieb, Arthur Dyevre

Nicolas Garneau, Piyush Ghai, Kimani Goddard

Ben Hagag, Nils Holzenberger

Karthik Kannapur, Daniel Katz, Ken Kawamura, Aykut Koc, Manolis Koubarakis

Ruta Liepina

Pauline McBride, Adam Meyers, Jelena Mitrović

Maria Navas-Loro, Joel Niklaus, Milda Norkute

Kartik Palani, Henrik Palmer Olsen

Georg Rehm

T.Y.S.S. Santosh, Vageesh Saxena, Madhavan Seshadri, Dan Simonson, Jerrold Soh, Alexandru Sotropa

Andrea Tagarelli, Dimitrios Tsarapatsanis

Josef Valvoda, Gijs Van Dijck, Jacob van de Kerkhof, Hellen van der Kroef

Jianqiang Wang, Hannes Westermann, Adam Wyner

Marcos Zampieri, Frederike Zufall

Keynote Talk

Privacy Protection, At What Cost?

Omri Ben-Shahar

University of Chicago Law School

Abstract: Data privacy protection is the dominant paradigm in the regulation of the digital economy. In this keynote lecture, University of Chicago Law Professor Omri Ben-Shahar challenges the adequacy of the data privacy paradigm. He argues that it fails to capture the most worrisome harms—what he calls ‘data pollution’—which are inflicted against public rather than private interests. He further demonstrates that privacy-fueled restrictions on valuable data technologies impose a large, mostly unrecognized, harms on society.

Bio: Omri Ben-Shahar earned his PhD in Economics and SJD from Harvard in 1995 and his BA and LLB from the Hebrew University in 1990. Before coming to Chicago, he was the Kirkland & Ellis Professor of Law and Economics at the University of Michigan. Prior to that, he taught at Tel-Aviv University, was a member of Israel’s Antitrust Court, and clerked at the Supreme Court of Israel. He teaches contracts, sales, trademark law, insurance law, consumer law, sales law, e-commerce, food law, law and economics, and game theory and the law. He writes primarily in the fields of contract law and consumer protection. He is the co-author of *Personalized Law: Different Rules for Different People* (Oxford 2021, with Ariel Porat) and *More Than You Wanted to Know: The Failure of Mandated Disclosure* (Princeton 2014, with Carl Schneider). Professor Ben-Shahar is the Kearney Director of the Coase-Sandor Institute for Law and Economics. He is also the Co-Reporter for the American Law Institute’s Restatement of Consumer Contracts.

Table of Contents

<i>LeGen: Complex Information Extraction from Legal sentences using Generative Models</i> Chaitra C R, Sankalp Kulkarni, Sai Rama Akash Varma Sagi, Shashank Pandey, Rohit Yalavarthy, Dipanjan Chakraborty and Prajna Devi Upadhyay	1
<i>Summarizing Long Regulatory Documents with a Multi-Step Pipeline</i> Mika Sie, Ruby Beek, Michiel Bots, Sjaak Brinkkemper and Albert Gatt	18
<i>Enhancing Legal Expertise in Large Language Models through Composite Model Integration: The Development and Evaluation of Law-Neo</i> Zhihao Liu, Yanzhen Zhu and Mengyuan Lu	33
<i>uOttawa at LegalLens-2024: Transformer-based Classification Experiments</i> Nima Meghdadi and Diana Inkpen	42
<i>Quebec Automobile Insurance Question-Answering With Retrieval-Augmented Generation</i> David Beauchemin, Richard Khoury and Zachary Gagnon	48
<i>Rethinking Legal Judgement Prediction in a Realistic Scenario in the Era of Large Language Models</i> Shubham Kumar Nigam, Aniket Deroy, Subhankar Maity and Arnab Bhattacharya	61
<i>The CLC-UKET Dataset: Benchmarking Case Outcome Prediction for the UK Employment Tribunal</i> Huiyuan Xie, Felix Steffek, Joana De Faria, Christine Carter and Jonathan Rutherford	81
<i>Information Extraction for Planning Court Cases</i> Drish Mali, Rubash Mali and Claire Barale	97
<i>Automated Anonymization of Parole Hearing Transcripts</i> Abed Itani, Wassiliki Siskou and Annette Hautli-Janisz	115
<i>Towards an Automated Pointwise Evaluation Metric for Generated Long-Form Legal Summaries</i> Shao Min Tan, Quentin Grail and Lee Quartey	129
<i>Enhancing Contract Negotiations with LLM-Based Legal Document Comparison</i> Savinay Narendra, Kaushal Shetty and Adwait Ratnaparkhi	143
<i>Attributed Question Answering for Preconditions in the Dutch Law</i> Felicia Redelaar, Romy Van Drie, Suzan Verberne and Maaïke De Boer	154
<i>Algorithm for Automatic Legislative Text Consolidation</i> Matias Etcheverry, Thibaud Real-del-Sarte and Pauline Chavallard	166
<i>Measuring the Groundedness of Legal Question-Answering Systems</i> Dietrich Trautmann, Natalia Ostapuk, Quentin Grail, Adrian Pol, Guglielmo Bonifazi, Shang Gao and Martin Gajek	176
<i>Transductive Legal Judgment Prediction Combining BERT Embeddings with Delaunay-Based GNNs</i> Hugo Attali and Nadi Tomeh	187
<i>Cross Examine: An Ensemble-based approach to leverage Large Language Models for Legal Text Ana- lytics</i> Saurav Chowdhury, Lipika Dey and Suyog Joshi	194
<i>LLMs to the Rescue: Explaining DSA Statements of Reason with Platform’s Terms of Services</i> Marco Aspromonte, Andrea Ferraris, Federico Galli and Giuseppe Contissa	205

<i>BLT: Can Large Language Models Handle Basic Legal Text?</i>	
Andrew Blair-Stanek, Nils Holzenberger and Benjamin Van Durme	216
<i>Multi-Property Multi-Label Documents Metadata Recommendation based on Encoder Embeddings</i>	
Nasredine Cheniki, Vidas Daudaravicius, Abdelfettah Feliachi, Didier Hardy and Marc Wilhelm Küster	233
<i>Comparative Study of Explainability Methods for Legal Outcome Prediction</i>	
Ieva Staliunaite, Josef Valvoda and Ken Satoh	243
<i>Bonafide at LegalLens 2024 Shared Task: Using Lightweight DeBERTa Based Encoder For Legal Violation Detection and Resolution</i>	
Shikha Bordia	259
<i>LAR-ECHR: A New Legal Argument Reasoning Task and Dataset for Cases of the European Court of Human Rights</i>	
Odysseas Chlapanis, Dimitris Galanis and Ion Androutsopoulos	267
<i>Gaps or Hallucinations? Scrutinizing Machine-Generated Legal Analysis for Fine-grained Text Evaluations</i>	
Abe Hou, William Jurayj, Nils Holzenberger, Andrew Blair-Stanek and Benjamin Van Durme	280
<i>Classify First, and Then Extract: Prompt Chaining Technique for Information Extraction</i>	
Alice Kwak, Clayton Morrison, Derek Bambauer and Mihai Surdeanu	303
<i>Augmenting Legal Decision Support Systems with LLM-based NLI for Analyzing Social Media Evidence</i>	
Ram Mohan Rao Kadiyala, Siddartha Pullakhandam, Kanwal Mehreen, Subhasya Tippareddy and Ashay Srivastava	318
<i>Empowering Air Travelers: A Chatbot for Canadian Air Passenger Rights</i>	
Maksym Taranukhin, Sahithya Ravi, Gabor Lukacs, Evangelos Milios and Vered Shwartz	326
<i>Enhancing Legal Violation Identification with LLMs and Deep Learning Techniques: Achievements in the LegalLens 2024 Competition</i>	
Nguyen Tan Minh, Duy Ngoc Mai, Le Xuan Bach, Nguyen Huu Dung, Pham Cong Minh, Ha Thanh Nguyen and Thi Hai Yen Vuong	336
<i>LegalLens 2024 Shared Task: Masala-chai Submission</i>	
Khalid Rajan and Royal Sequiera	346
<i>Semantists at LegalLens-2024: Data-efficient Training of LLM's for Legal Violation Identification</i>	
Kanagasabai Rajaraman and Hariram Veeramani	355
<i>LegalLens Shared Task 2024: Legal Violation Identification in Unstructured Text</i>	
Ben Hagag, Gil Gil Semo, Dor Bernsohn, Liav Harpaz, Pashootan Vaezipoor, Rohit Saha, Kyryl Truskovskiy and Gerasimos Spanakis	361
<i>DeBERTa Beats Behemoths: A Comparative Analysis of Fine-Tuning, Prompting, and PEFT Approaches on LegalLensNER</i>	
Hanh Thi Hong Tran, Nishan Chatterjee, Senja Pollak and Antoine Doucet	371
<i>LexSumm and LexT5: Benchmarking and Modeling Legal Summarization Tasks in English</i>	
Santosh T.Y.S.S, Cornelius Weiss and Matthias Grabmair	381
<i>Towards Supporting Legal Argumentation with NLP: Is More Data Really All You Need?</i>	
Santosh T.Y.S.S, Kevin Ashley, Katie Atkinson and Matthias Grabmair	404

Program

Saturday, November 16, 2024

09:00 - 09:15 *Workshop Opening*

09:10 - 10:30 *Session 1*

Summarizing Long Regulatory Documents with a Multi-Step Pipeline

Mika Sie, Ruby Beek, Michiel Bots, Sjaak Brinkkemper and Albert Gatt

Towards an Automated Pointwise Evaluation Metric for Generated Long-Form Legal Summaries

Shao Min Tan, Quentin Grail and Lee Quartey

Cross Examine: An Ensemble-based approach to leverage Large Language Models for Legal Text Analytics

Saurav Chowdhury, Lipika Dey and Suyog Joshi

LexSumm and LexT5: Benchmarking and Modeling Legal Summarization Tasks in English

Santosh T.Y.S.S, Cornelius Weiss and Matthias Grabmair

Algorithm for Automatic Legislative Text Consolidation

Matias Etcheverry, Thibaud Real-del-Sarte and Pauline Chavallard

LeGen: Complex Information Extraction from Legal sentences using Generative Models

Chaitra C R, Sankalp Kulkarni, Sai Rama Akash Varma Sagi, Shashank Pandey, Rohit Yalavarthy, Dipanjan Chakraborty and Prajna Devi Upadhyay

Information Extraction for Planning Court Cases

Drish Mali, Rubash Mali and Claire Barale

Automated Anonymization of Parole Hearing Transcripts

Abed Itani, Wassiliki Siskou and Annette Hautli-Janisz

BLT: Can Large Language Models Handle Basic Legal Text?

Andrew Blair-Stanek, Nils Holzenberger and Benjamin Van Durme

Classify First, and Then Extract: Prompt Chaining Technique for Information Extraction

Alice Kwak, Clayton Morrison, Derek Bambauer and Mihai Surdeanu

Saturday, November 16, 2024 (continued)

HiCuLR: Hierarchical Curriculum Learning for Rhetorical Role Labeling of Legal Documents

Santosh T.Y.S.S, Apolline Isaia, Shiyu Hong and Matthias Grabmair

10:30 - 11:00 *Break*

11:00 - 12:10 *Session 2*

Rethinking Legal Judgement Prediction in a Realistic Scenario in the Era of Large Language Models

Shubham Kumar Nigam, Aniket Deroy, Subhankar Maity and Arnab Bhattacharya

The CLC-UKET Dataset: Benchmarking Case Outcome Prediction for the UK Employment Tribunal

Huiyuan Xie, Felix Steffek, Joana De Faria, Christine Carter and Jonathan Rutherford

Transductive Legal Judgment Prediction Combining BERT Embeddings with Delaunay-Based GNNs

Hugo Attali and Nadi Tomeh

Comparative Study of Explainability Methods for Legal Outcome Prediction

Ieva Staliunaite, Josef Valvoda and Ken Satoh

Incorporating Precedents for Legal Judgement Prediction on European Court of Human Rights Cases

Santosh T.Y.S.S, Mohamed Hesham Elganayni, Stanisław Sójka and Matthias Grabmair

The Craft of Selective Prediction: Towards Reliable Case Outcome Classification - An Empirical Study on European Court of Human Rights Cases

Santosh T.Y.S.S, Irtiza Chowdhury, Shanshan Xu and Matthias Grabmair

Quebec Automobile Insurance Question-Answering With Retrieval-Augmented Generation

David Beauchemin, Richard Khoury and Zachary Gagnon

Attributed Question Answering for Preconditions in the Dutch Law

Felicia Redelaar, Romy Van Drie, Suzan Verberne and Maaïke De Boer

Measuring the Groundedness of Legal Question-Answering Systems

Dietrich Trautmann, Natalia Ostapuk, Quentin Grail, Adrian Pol, Guglielmo Bonifazi, Shang Gao and Martin Gajek

Saturday, November 16, 2024 (continued)

12:10 - 14:00 *Lunch and In-Person Poster Session*

uOttawa at LegalLens-2024: Transformer-based Classification Experiments

Nima Meghdadi and Diana Inkpen

Bonafide at LegalLens 2024 Shared Task: Using Lightweight DeBERTa Based Encoder For Legal Violation Detection and Resolution

Shikha Bordia

Augmenting Legal Decision Support Systems with LLM-based NLI for Analyzing Social Media Evidence

Ram Mohan Rao Kadiyala, Siddhartha Pullakhandam, Kanwal Mehreen, Subhasya Tippareddy and Ashay Srivastava

Enhancing Legal Violation Identification with LLMs and Deep Learning Techniques: Achievements in the LegalLens 2024 Competition

Nguyen Tan Minh, Duy Ngoc Mai, Le Xuan Bach, Nguyen Huu Dung, Pham Cong Minh, Ha Thanh Nguyen and Thi Hai Yen Vuong

LegalLens 2024 Shared Task: Masala-chai Submission

Khalid Rajan and Royal Sequiera

Semantists at LegalLens-2024: Data-efficient Training of LLM's for Legal Violation Identification

Kanagasabai Rajaraman and Hariram Veeramani

DeBERTa Beats Behemoths: A Comparative Analysis of Fine-Tuning, Prompting, and PEFT Approaches on LegalLensNER

Hanh Thi Hong Tran, Nishan Chatterjee, Senja Pollak and Antoine Doucet

14:00 - 15:00 *Keynote - Privacy Protection, At What Cost? - Omri Ben-Shahar (University of Chicago Law School)*

15:00 - 15:30 *Shared Task*

LegalLens Shared Task 2024: Legal Violation Identification in Unstructured Text

Ben Hagag, Gil Gil Semo, Dor Bernsohn, Liav Harpaz, Pashootan Vaezipoor, Rohit Saha, Kyril Truskovskiy and Gerasimos Spanakis

15:30 - 16:00 *Break*

Saturday, November 16, 2024 (continued)

16:00 - 17:30 *Session 4*

LLMs to the Rescue: Explaining DSA Statements of Reason with Platform's Terms of Services

Marco Aspromonte, Andrea Ferraris, Federico Galli and Giuseppe Contissa

Enhancing Contract Negotiations with LLM-Based Legal Document Comparison

Savinay Narendra, Kaushal Shetty and Adwait Ratnaparkhi

Multi-Property Multi-Label Documents Metadata Recommendation based on Encoder Embeddings

Nasredine Cheniki, Vidas Daudaravicius, Abdelfettah Feliachi, Didier Hardy and Marc Wilhelm Küster

CLERC: A Dataset for U. S. Legal Case Retrieval and Retrieval-Augmented Analysis Generation

Abe Hou, Orion Weller, Guanghui Qin, Eugene Yang, Dawn Lawrie, Nils Holzenberger, Andrew Blair-Stanek and Benjamin Van Durme

Empowering Air Travelers: A Chatbot for Canadian Air Passenger Rights

Maksym Taranukhin, Sahithya Ravi, Gabor Lukacs, Evangelos Milios and Vered Shwartz

The impact of formulaic language in the Court of Justice of the European Union on the performance of lexical and dense retrieval methods

Larissa Mori, Carlos Sousa De Oliveira, Yuehwern Yih and Mario Ventresca

Gaps or Hallucinations? Scrutinizing Machine-Generated Legal Analysis for Fine-grained Text Evaluations

Abe Hou, William Jurayj, Nils Holzenberger, Andrew Blair-Stanek and Benjamin Van Durme

How Many Van Goghs Does It Take to Van Gogh? Finding the Imitation Threshold

Sahil Verma, Royi Rassin, Arnav Mohanty Das, Gantavya Bhatt, Preethi Seshadri, Chirag Shah, Jeff Bilmes, Hannaneh Hajishirzi and Yanai Elazar

Towards Supporting Legal Argumentation with NLP: Is More Data Really All You Need?

Santosh T.Y.S.S, Kevin Ashley, Katie Atkinson and Matthias Grabmair

Misinformation with Legal Consequences (MisLC): A New Task Towards Harnessing Societal Harm of Misinformation

Chu Fei Luo, Radin Shayanfar, Rohan Bhambhoria, Samuel Dahan and Xiaodan Zhu

Saturday, November 16, 2024 (continued)

LAR-ECHR: A New Legal Argument Reasoning Task and Dataset for Cases of the European Court of Human Rights

Odysseas Chlapanis, Dimitris Galanis and Ion Androutsopoulos

Developing a Pragmatic Benchmark for Assessing Korean Legal Language Understanding in Large Language Models

Yeeun Kim, Young Rok Choi, Eun Kyung Choi, Jin Hwan Choi, Hai Jin Park and Wonseok Hwang

Enhancing Legal Expertise in Large Language Models through Composite Model Integration: The Development and Evaluation of Law-Neo

Zhihao Liu, Yanzhen Zhu and Mengyuan Lu

17:30 - 17:40 *Closing Remarks & Best Presentation Award*