

# Natural Legal Language Processing (NLLP) Workshop 2021

**Nikos Aletras** - University of Sheffield

**Ion Androutsopoulos** - Athens University of Economics and Business

**Leslie Barrett** - Bloomberg Law

**Cătălina Goanță** - Maastricht University

**Daniel Preoțiuc-Pietro** - Bloomberg

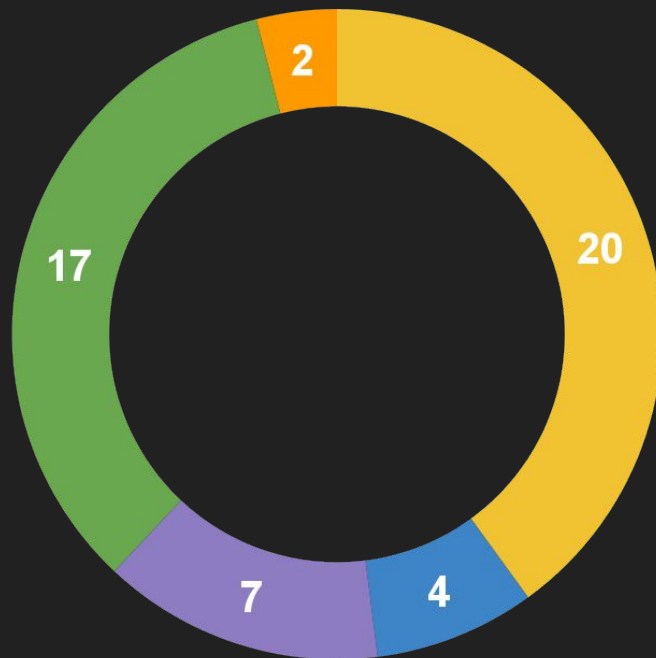
# Aims

- Bring together **Natural Language Processing (NLP)**, **Machine Learning (ML)** and **Law** communities
- **Foster collaboration** between these areas
- Discuss **applications in the legal domain**
- **Formulate targets and priorities** for computational approaches to legal text
- **Identify challenges** specific to the legal domain

# Paper Types

- New **Applications** of NLP to legal tasks
- **Experimental**
  - Studies adapting NLP methods to legal documents
- **Tasks**
  - Description of new legal tasks for NLP
- **Resources**
  - New datasets than can be publicly released and reused
- **Demos**
  - Description of systems using NLP for legal text
- **Industrial** Research
- Interdisciplinary **position papers**

# Paper Stats



● Not accepted ● Non-Archival ● Short ● Long ● Findings

# Program Committee

Tomaso Agnoloni — [Institute of Legal Information Theory and Technologies](#)

Elliott Ash — [ETH Zurich](#)

Breck Baldwin — [Columbia University](#)

Andrew Blair-Stanek — [Johns Hopkins University](#)

Lukasz Borchmann — [Applica.ai](#)

Ilias Chalkidis — [University of Copenhagen](#)

Rajarathnam Chandramouli — [Stevens Institute of Technology](#)

Daniel Chen — [Toulouse School of Economics](#)

Marina Danilevsky — [IBM Research](#)

Tony Davis — [Southern Oregon University](#)

Stefania Degaetano-Ortlieb — [Saarland University](#)

Luigi Di Caro — [University of Turin](#)

Fabiana Di Porto — [University of Salento](#)

Kasper Drazewski — [The European Consumer Organization](#)

Arthur Dyevre — [Katholieke Universiteit Leuven](#)

Emmanouil Fergadiotis — [AUEB and NCSR “Demokritos”](#)

Frank Giaoui — [Columbia Law School](#)

Matthias Grabmair — [Technical University of Munich](#)

Nils Holzenberger — [Johns Hopkins University](#)

Dan Katz — [Illinois Tech - Chicago Kent College of Law](#)

Ilan Kernerman — [K Dictionaries](#)

Manolis Kourbarakis — [University of Athens](#)

Vasileios Lampos — [University College London](#)

Junyi Jessie Li — [University of Texas at Austin](#)

Ruta Liepina — [Maastricht University](#)

How Khang Lim — [Singapore Management University](#)

Makis Malakasiotis — [AUEB and NCSR “Demokritos”](#)

Adam Meyers — [New York University](#)

Jelena Mitrovic — [University of Passau](#)

Paulo Quaresma — [Universidade de Évora](#)

Georg Rehm — [DFKI](#)

George Sanchez — [Thomson Reuters](#)

Thibault Schrepel — [Vrije Universiteit](#)

Mathias Siems — [European University Institute](#)

Dan Simonson — [Blackboiler](#)

Jerrold Soh — [Singapore Management University](#)

Gerasimos Spanakis — [Maastricht University](#)

Andrea Tagarelli — [University of Calabria](#)

Dimitrios Tsarapatsanis — [University of York](#)

Mihaela Vela — [Saarland University](#)

Adam Wyner — [Swansea University](#)

Marcos Zampieri — [Rochester Institute of Technology](#)

We would like to thank our reviewers - Each paper received 3+ reviews

# Schedule

08:15 - 09:00 **Invited Speaker: John Armour**

09:00 - 10:30 **Session 1: Data Sets**

10:45 - 12:10 **Session 2: Prediction I**

12:10 - 13:00 **Virtual Town Hall & Lunch**

13:00 - 13:45 **Invited Speaker: Sylvie Delacroix**

13:45 - 14:30 **Session 3: Retrieval**

14:45 - 16:10 **Session 4: Information Extraction**

16:25 - 17:20 **Session 5: Generation**

17:20 - 18:00 **Session 6: Prediction II**

# Workshop Materials



Zoom direct meeting link:

<https://maastrichtuniversity.zoom.us/j/93586782402>

YouTube channel:

<https://www.youtube.com/channel/UCqqT9g-V0lsgHloVF6S75dw>



1. Recordings
  - Available within a week after the workshop on Youtube
2. Proceedings
  - ACL Anthology: <https://aclanthology.org/events/emnlp-2021/#2021-nllp-1>

# Chat / Q&A

## Zoom chat

- Only available during the live session





# Talks

- Monthly talk series
- In-depth presentations of recent research at the intersection of law and NLP
- Format: 30 min presentation + discussant from the complementary discipline
- Monthly on Fridays at 8am PST/11am EST/16 UK/17 CET time.
- Schedule and past talks: <https://nllpw.org/talks/>

**NLLP TALKS**



# YouTube

- Monthly NLLP talks
- NLLP 2020 workshop
- NLLP 2021 workshop will also be posted on the channel

The screenshot shows the YouTube channel page for 'Natural Legal Language Processing Workshop'. The channel has 53 subscribers and is marked as 'SUBSCRIBED'. The navigation menu includes HOME, VIDEOS, PLAYLISTS, CHANNELS, and ABOUT. The 'Uploads' section is active, showing three video thumbnails:

- NLLP Talks #3 (Josef Valvoda)**: 101 views • 4 months ago. Video duration: 58:40.
- NLLP Talks #2 (Andrea Galassi & Kasper Drazewski)**: 65 views • 5 months ago. Video duration: 1:07:30. The thumbnail includes a list of topics: 'A broad legal NLP experimentation project', 'Three generations of young scholars', 'EU, Academia and Pragmatism, Belgium, Polish Academy of Sciences', 'Collection of earlier research in online Justice of Service', 'Light compliance testing of privacy policies', 'Providing justifications for declared potential unfairness', 'Linguistic transparency analysis', and 'Multilingualism (following UD in multiple languages)'.
- NLLP Talks #1 (Ilias Chalkidis)**: 109 views • 6 months ago. Video duration: 1:04:21. The thumbnail includes the text: 'NLLP TALKS', 'PARAGRAPHS-LEVEL NARRATIVE EXTRACTION THROUGH REGULARIZATION: A CASE STUDY ON EUROPEAN COURT OF HUMAN RIGHTS CASES', 'PRESENTATION BY ILIAS CHALKIDIS (LATEINUS UNIVERSITY OF ECONOMICS AND BUSINESS)', and 'MODERATION BY: SARA VAN DER BREEK (BRASCHTICHT LIFE & TECH LAB)'. The event date is 'FRIDAY, 16 APRIL 2021'.

Below the video thumbnails, there are three playlist cards:

- Session 3: Data Sets**: A Dataset for Statutory Reasoning in Tax Law Establishment and Question Answering. Mentions: Alja Hübenberger, Andrew Blah-Sternik, Benjamin Van Durme. Source: MIVCC: A Corpus of Maltese Criminal Cases.
- Session 2: NLP Applications**: Toward Domain-Guided Contrastive Summarization of Privacy Policies. Mentions: Marika Koyamanchi, Mihai Eftene, Srinivasan Parthasarathy. Source: Entity in Legal Language.
- Session 4: System Demos**: Quick Check: A Legal Research Recommendation System. Mentions: Marthe Thomas, Thomas Vassil, Xin Shui, Wenkai Luo, George Sanchez, Paras Gethia, Don Teo, Kishka Madan, Tonya Guella.

# Messaging

- Slack
  - <https://nlp.slack.com/>
  - [Direct join link](#)



- Google Groups - Mailing list
  - <https://groups.google.com/d/forum/nlp>

# Website

- Website: <https://nllpw.org/>
- Resources: <https://nllpw.org/resources/>
  - Relevant publications
  - Data sets
  - Models
  - Related events

# Twitter

- Hashtags
  - [#nllp](#) - Official workshop hashtag
  - [#EMNLP2021](#) - Official conference hashtag
  - [#legaltech](#) - Computational Legal hashtag
  - [#nlproc](#) - Natural Language Processing hashtag
- [@emnlpmeeting](#) - Official conference account
- [@nllpworkshop](#) - Official workshop account

# Invited Speaker: John Armour

08:15 - 09:00 **Access to Caselaw Data in the UK:  
Constraints and Future Prospects**

- In this talk, we will consider the legal, regulatory and practical constraints on access to caselaw data in the UK for use in legal NLP applications, contrasting these with the position in other leading jurisdictions. We will analyse how the legal constraints can be managed and provide an overview of the future prospects for data access.



# Session 1: Data Sets

09:00 - 09:15      A Corpus for Multilingual Analysis of Online Terms of Service

*Kasper Drawzeski, Andrea Galassi, Agnieszka Jablonowska, Francesca Lagioia, Marco Lippi, Hans Micklitz, Giovanni Sartor, Giacomo Tagiuri, Paolo Torroni*

09:15 - 09:30      Named Entity Recognition in the Romanian Legal Domain

*Vasile Pais, Maria Mitrofan, Carol Luca Gasan, Vlad Coneschi, Alexandru Ianov*

09:30 - 09:45      The Power of Legislatures in Hungary - A Text Reuse Analysis

*Miklós Sebők, Anna Székely, István Járay*

09:45 - 10:00      Swiss-Judgment-Prediction: A Multilingual Legal Judgment Prediction Benchmark

*Joel Niklaus, Ilias Chalkidis, Matthias Stürmer*

10:00 - 10:15      LexGLUE: A Benchmark Dataset for Legal Language Understanding in English

*Ilias Chalkidis, Abhik Jana, Dirk Hartung, Michael Bommarito, Ion Androutsopoulos, Daniel Martin Katz, Nikolaos Aletras*

10:15 - 10:30      'Just What do You Think You're Doing, Dave?' A Checklist for Responsible Data Use in NLP

*Anna Rogers, Tim Baldwin, Kobi Leins*

# Session 2: Prediction I

10:45 - 11:00 Automated Extraction of Sentencing Decisions from Court Cases in the Hebrew Language

*Mohr Wenger, Tom Kalir, Noga Berger, Carmit Chalamish, Renana Keydar, Gabriel Stanovsky*

11:00 - 11:15 A Multilingual Approach to Identify and Classify Exceptional Measures against COVID-19

*Georgios Tziafas, Eugenie de Saint-Phalle, Wietse de Vries, Clara Egger, Tommaso Caselli*

11:15 - 11:30 Multi-granular Legal Topic Classification on Greek Legislation

*Christos Papaloukas, Ilias Chalkidis, Konstantinos Athinaios, Despina Pantazi, Manolis Koubarakis*

11:30 - 11:40 Machine Extraction of Tax Laws from Legislative Texts

*Elliott Ash, Malka Guillot, Luyang Han*

11:40 - 11:50 jurBERT: A Romanian BERT Model for Legal Judgement Prediction

*Mihai Masala, Radu Cristian Alexandru Iacob, Ana Sabina Uban, Marina Cidota, Horia Velicu, Traian Rebedea, Marius Popescu*

11:50 - 12:00 JuriBERT: A Masked-Language Model Adaptation for French Legal Text

*Stella Douka, Hadi Abdine, Michalis Vazirgiannis, Rajaa El Hamdani, David Restrepo Amariles*

12:00 - 12:10 Few-shot and Zero-shot Approaches to Legal Text Classification: A Case Study in the Financial Sector

*Rajdeep Sarkar, Atul Kr. Ojha, Jay Megaro, John Mariano, Vall Herard, John P. McCrae*



# NLLP Virtual Town Hall & Lunch

Let's get together and discuss:

- Highlights
- Lowlights - How can we make NLLP better?
- Future directions
  - Areas to focus?
  - Shared tasks?
- Seeking PC members, especially from
  - Law
  - Asia-Pacific

Bring your own lunch

# Invited Speaker: Sylvie Delacroix

## 13:00 - 13:45 **Data Trusts as a Bottom-up Empowerment Tool**

This presentation will focus on data trusts as a bottom-up empowerment tool. It will proceed from an analysis of the particular type of vulnerability concomitant with our 'leaking' data on a daily basis, to argue that data ownership is both unlikely and inadequate as an answer to the problems at stake. There are three key problems that bottom-up data trusts seek to address:

- Lack of mechanisms to empower groups, not just individuals
- We can (and should) do better than 'make belief' consents
- We can (and should) do better when it comes to intelligent data sharing.



# Session 3: Retrieval

13:45 - 14:00      AutoLAW: Augmented Legal Reasoning through Legal Precedent Prediction

*Robert Zev Mahari*

14:00 - 14:10      A Free Format Legal Question Answering System

*Soha Khazaeli, Janardhana Punuru, Chad Morris, Sanjay Sharma, Bert Staub, Michael Cole, Sunny Chiu-Webster, Dhruv Sakalley*

14:10 - 14:20      Searching for Legal Documents at Paragraph Level: Automating Label Generation and Use of an Extended Attention Mask for Boosting Neural Models of Semantic Similarity

*Li Tang and Simon Clematide*

14:20 - 14:30      GerDaLIR: A German Dataset for Legal Information Retrieval

*Marco Wrzalik and Dirk Krechel*

# Session 4: Information Extraction

14:45 - 15:00 SPaR.txt, a Cheap Shallow Parsing Approach for Regulatory Texts

Ruben Kruiper, Ioannis Konstas, Alasdair Gray, Farhad Sadeghineko, Richard Watson, Bimal Kumar

15:00 - 15:15 Capturing Logical Structure of Visually Structured Documents with Multimodal Transition Parser

*Yuta Koreeda and Christopher Manning*

15:15 - 15:30 Legal Terminology Extraction with the Termolator

*Nhi Pham, Lachlan Pham, Adam Meyers*

15:30 - 15:45 Supervised Identification of Participant Slots in Contracts

*Dan Simonson*

15:45 - 16:00 ContractNLI: A Dataset for Document-level Natural Language Inference for Contracts

*Yuta Koreeda and Christopher Manning*

16:00 - 16:10 *Named Entity Recognition in Historic Legal Text: A Transformer and State Machine Ensemble Method*

*Fernando Trias, Hongming Wang, Sylvain Jaume, Stratos Idreos*

# Session 5: Generation

16:25 - 16:40 Summarization of German Court Rulings

*Ingo Glaser, Sebastian Moser, Florian Matthes*

16:40 - 16:55 Privacy Policy Question Answering Assistant A Query-Guided Extractive Summarization Approach

*Moniba Keymanesh, Micha Elsner, Srinivasan Parthasarathy*

16:55 - 17:10 Learning from Limited Labels for Long Legal Dialogue

*Jenny Hong, Derek Chong, Christopher Manning*

17:10 - 17:20 Automating Claim Construction in Patent Applications: The CMUmine Dataset

*Ozan Tonguz, Yiwei Qin, Yimeng Gu, Hyun Moon*

# Session 6: Prediction II

17:20 - 17:30      Effectively Leveraging BERT for Legal Document Classification  
*Nut Limsopatham*

17:30 - 17:45      Semi-automatic Triage of Requests for Free Legal Assistance  
*Meladel Mistica, Jey Han Lau, Brayden Merrifield, Kate Fazio, Timothy Baldwin*

17:45 - 18:00      Automatic Resolution of Domain Name Disputes  
*Wayan Vihikan, Meladel Mistica, Inbar Levy, Andrew Christie, Timothy Baldwin*