LREC-COLING 2024

Legal and Ethical Issues in Human Language Technologies 2024 (LEGAL2024)

Workshop Proceedings

Editors
Ingo Siegert and Khalid Choukri

20 May, 2024
Torino, Italia

Proceedings of LEGAL2024: Workshop on Legal and Ethical Issues in Human Language Technologies @LREC-COLING-2024

ISBN 978-2-493814-21-0
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Preface

The year 2023 witnessed extensive discussions revolving around Artificial Intelligence (AI) and Large Language Models (LLM), marking a significant era in technological advancements. These discussions shed light on the unprecedented collection and utilization of data required by such technologies, often owned by stakeholders not directly involved in their development. Repackaging and repurposing these vast language datasets for AI and LLM endeavors become imperative, despite the intangible nature of language data, as they are subject to legal constraints.

In recent years, substantial efforts have been dedicated to adapting legal frameworks to technological advancements while considering the interests of diverse stakeholders. However, the strict consideration of legal aspects poses additional questions beyond mere recording technology and participant consent, constituting several key elements that warrant attention.

The LREC Workshop on Legal and Ethical Issues in Human Language Technologies 2024, held on Monday, June 20, 2024, as part of the LREC 2024 Conference, aims to delve into these crucial aspects. This workshop serves as a platform to explore the intricate interactions between legal and technical dimensions of data collection, processing, and distribution, particularly focusing on text crawling, speech and voice recordings, and the implications of text and speech data mining exceptions introduced by legislative bodies. Furthermore, it examines the compatibility of legal requirements for data collection and processing, as mandated by regulations like GDPR, alongside the technical feasibility of various anonymization and pseudonymization techniques.

A highlight of the workshop includes an invited talk by Jennifer Williams from the University of Southampton, offering insights into "AI Regulation Perspectives from the UK". This talk promises to enrich discussions by providing a comprehensive overview of AI regulation in the UK context. Furthermore, the workshop featured 11 accepted papers covering a range of topics. These included the use of LLMs in Finnish higher education, implications of regulations in the US super election year, intellectual property rights, annotating hate speech data, selling personal information, cultural heritage and data collection, and a comparison of voice user usage between Germany and Finland. These papers provided valuable insights into the legal and ethical challenges facing language technologies.

The workshop also delves into broader issues encompassing ethics, morality, and trust, exploring their interplay with data collection and distribution. It aims to foster dialogue between technology and legal experts, addressing current legal and ethical challenges in the Human Language Technology sector.

This volume encapsulates the proceedings of the LREC Workshop on Legal and Ethical Issues in Human Language Technologies 2024, documenting valuable insights and discussions shared during the event. We extend our gratitude to our keynote speakers and authors for their contributions, as well as to the Program Committee for their diligent review efforts.

Ingo Siegert, Khalid Choukri & Pawel Kamocki, April 2024

Organizing Committee

Ingo Siegert, Otto-von-Guericke-Universität Magdeburg, Germany
Khalid Choukri, ELRA/ELDA, France
Pawel Kamocki, IDS Mannheim, Germany
Kossay Talmoudi, ELDA, France

Table of Contents

LEGAL2024 Program

09:00 - 09:15 **Opening Session: Welcome by Workshop Chairs**

09:15 - 09:30 Participant and Organizer Introduction

09:30 - 10:30 Invited Talk: AI Regulation Perspectives from the UK (Jennifer Williams, University of Southampton)

10:30 - 11:00 **Coffee Break**

11:00 - 13:00 **Session I: Legal Frameworks and Ethical Considerations**

Compliance by Design Methodologies in the Legal Governance Schemes of European Data Spaces *(Kossay Talmoudi, Khalid Choukri, and Isabelle Gavanon)*

A Legal Framework for Natural Language Model Training in Portugal *(Ruben Almeida and Evelin Amorim)*

Intellectual property rights at the training, development and generation stages of Large Language Models *(Christin Kirchhübel and Georgina Brown)*

Ethical Issues in Language Resources and Language Technology – New Challenges, New Perspectives *(Pawel Kamocki and Andreas Witt)*

13:00 - 14:00 **Lunch Break**

14:00 - 16:00 **Session II: Considerations and Implications of AI**

Legal and Ethical Considerations that Hinder the Use of LLMs in a Finnish Institution of Higher Education *(Mika Hämäläinen)*

Implications of Regulations on Large Generative AI Models in the Super-Election Year and the Impact on Disinformation *(Vera Schmitt, Jakob Tesch, Eva Lopez, Tim Polzehl, Aljoscha Burchardt, Konstanze Neumann, Salar Mohtaj and Sebastian Möller)*

Selling Personal Information: Data Brokers and the Limits of US Regulation (Denise DiPersio)

What can I do with this data point? Towards modeling legal and ethical aspects of linguistic data collection and (re)use as a process *(Annett Jorschick, Paul T. Schrader and Hendrik Buschmeier)*

16:00 - 16:30 **Coffee Break**

16:30 - 17:30 **Session III: Applications and User Perspective**

Data-Envelopes for Cultural Heritage: Going beyond Datasheets *(Maria Eskevich and Mrinalini Luthra)*

Emotional Toll and Coping Strategies: Navigating the Effects of Annotating Hate Speech Data *(Maryam M. AlEmadi and Wajdi Zaghouani)*

User Perspective on Anonymity in Voice Assistants – A comparison between Germany and Finland *(Ingo Siegert, Silas Rech, Matthias Haase and Tom Bäckström)*

17:30 - 18:00 **Wrap-Up of the Workshop and Closing Ceremony**