Legal and Ethical Issues in Human Language Technologies and Multilingual De-Identification of Sensitive Language Resources (LEGAL-MDLR 2022)

Held at LREC 2022

Marseille, France
20 June 2022

Editors:

Mickael Rigault
Victoria Arranz
Ingo Siegert

ISBN: 978-1-7138-6126-3

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2022) by European Language Resources Association (ELRA)
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-NC. (https://creativecommons.org/licenses/by-nc/4.0/)

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Keynote Speech - Major Developments in the Legal Framework Concerning Language Resources
Pawel Kamocki 1

Sentiment Analysis and Topic Modeling for Public Perceptions of Air Travel: COVID Issues and Policy Amendments
Avery Field, Aparna Varde and Pankaj Lal 2

Data Protection, Privacy and US Regulation
Denise DiPersio 9

Pseudonymisation of Speech Data as an Alternative Approach to GDPR Compliance
Pawel Kamocki and Ingo Siegert 17

Categorizing Legal Features in a Metadata-Oriented Task: Defining the Conditions of Use
Mickaël Rigault, Victoria Arranz, Valérie Mapelli, Penny Labropoulou and Stelios Piperidis . . . 22

About Migration Flows and Sentiment Analysis on Twitter data: Building the Bridge between Technical and Legal Approaches to Data Protection
Thilo Gottschalk and Francesca Pichierri 27

Transparency and Explainability of a Machine Learning Model in the Context of Human Resource Management
Sebastien Delecraz, Loukman Eltarr and Olivier Oullier 38

Public Interactions with Voice Assistant – Discussion of Different One-Shot Solutions to Preserve Speaker Privacy
Ingo Siegert, Yamini Sinha, Gino Winkelmann, Oliver Jokisch and Andreas Wendemuth 44

Keynote Speech - Voice Anonymization and the GDPR
Brij Mohan Lal Srivastava 48

Cross-Clinic De-Identification of Swedish Electronic Health Records: Nuances and Caveats
Olle Bridal, Thomas Vakili and Marina Santini 49

Generating Realistic Synthetic Curricula Vitae for Machine Learning Applications under Differential Privacy
Andrea Bruera, Francesco Aldà and Francesco Di Cerbo 53

MAPA Project: Ready-to-Go Open-Source Datasets and Deep Learning Technology to Remove Identifying Information from Text Documents
Victoria Arranz, Khalid Choukri, Montse Cuadros, Aitor García Pablos, Lucie Gianola, Cyril Grouin, Manuel Herranz, Patrick Paroubek and Pierre Zweigenbaum 64

PriPA: A Tool for Privacy-Preserving Analytics of Linguistic Data
Jeremie Clos, Emma McClaughlin, Pepita Barnard, Elena Nichele, Dawn Knight, Derek McAuley and Svenja Adolphs 73

Legal and Ethical Challenges in Recording Air Traffic Control Speech
Mickaël Rigault, Claudia Cevenini, Khalid Choukri, Martin Kocour, Karel Veselý, Igor Szoke, Petr Motlicek, Juan Pablo Zuluaga-Gomez, Alexander Blatt, Dietrich Klakow, Allan Tart, Pavel Kolčárek and Jan Černocký 79

It is not Dance, is Data: Gearing Ethical Circulation of Intangible Cultural Heritage Practices in the Digital Space
Jorge Yánez and Amel Fraisse 84