

PrivateNLP 2025

**The Sixth Workshop on Privacy in Natural Language
Processing**

Proceedings of the Workshop

April 4, 2025

©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 979-8-89176-246-6

Introduction

Welcome to the Sixth Workshop on Privacy in Natural Language Processing. Co-located with NAA-CL 2025 in Albuquerque (NM), USA, the workshop is scheduled for April 4, 2025. To facilitate the participation of the global NLP community, we continue running the workshop in a hybrid format.

Privacy-preserving language data processing has become essential in the age of Large Language Models (LLMs) where access to vast amounts of data can provide gains over tuned algorithms. A large proportion of user-contributed data comes from natural language e.g., text transcriptions from voice assistants. It is therefore important to curate NLP datasets while preserving the privacy of the users whose data is collected, and train ML models that only retain non-identifying user data. The workshop brings together practitioners and researchers from academia and industry to discuss the challenges and approaches to designing, building, verifying, and testing privacy preserving systems in the context of Natural Language Processing.

Our agenda features a keynote speech, hybrid talk sessions both for long and short papers, and a poster session. This year we received 13 submissions. We accepted 9 submissions after a thorough peer-review. One accepted submissions has been withdrawn by the authors.

We would like to deeply thank to all the authors, committee members, keynote speaker, and participants to help us make this research community grow both in quantity and quality.

Workshop Chairs

Organizing Committee

Program Chairs

Ivan Habernal, Ruhr-University Bochum, Germany

Sepideh Ghanavati, University of Maine, United States

Vijayanta Jain, University of Maine, United States

Timour Igamberdiev, University of Vienna, Austria

Shomir Wilson, Pennsylvania State University, United States

Program Committee

Reviewers

Gergely Acs, Technical University of Budapest
Stefan Arnold, Friedrich-Alexander-Universität
Andrea Atzeni, Polytechnic Institute of Turin
Travis Breaux, Carnegie Mellon University
Christos Dimitrakakis, Université de Neuchâtel, University of Oslo and Chalmers University
Natasha Fernandes, Macquarie University
James Flemings, University of Southern California
Pierre Lison, Norwegian Computing Center
Christina Lohr, Universität Leipzig
Eugenio Martínez-Cámara, Universidad de Jaén
Stephen Meisenbacher, Technische Universität München
Isar Nejadgholi, National Research Council Canada and University of Ottawa
Sebastian Ochs, Technische Universität Darmstadt
Sai Teja Peddinti, Google
Lizhen Qu, Monash University
Afsaneh Razi, Drexel University
Peter Story, Clark University
David Sánchez, Universitat Rovira i Virgili
Ruyu Zhou, University of Notre Dame

Table of Contents

<i>TUNI: A Textual Unimodal Detector for Identity Inference in CLIP Models</i> Songze Li, Ruoxi Cheng and Xiaojun Jia	1
<i>TAROT: Task-Oriented Authorship Obfuscation Using Policy Optimization Methods</i> Gabriel Loiseau, Damien Sileo, Damien Riquet, Maxime Meyer and Marc Tommasi	14
<i>Balancing Privacy and Utility in Personal LLM Writing Tasks: An Automated Pipeline for Evaluating Anonymizations</i> Stefan Pasch and Min Chul Cha	32
<i>Named Entity Inference Attacks on Clinical LLMs: Exploring Privacy Risks and the Impact of Mitigation Strategies</i> Adam Sutton, Xi Bai, Kawsar Noor, Thomas Searle and Richard Dobson	42
<i>Inspecting the Representation Manifold of Differentially-Private Text</i> Stefan Arnold	53
<i>Beyond Reconstruction: Generating Privacy-Preserving Clinical Letters</i> Libo Ren, Samuel Belkadi, Lifeng Han, Warren Del-Pinto and Goran Nenadic	60
<i>Beyond De-Identification: A Structured Approach for Defining and Detecting Indirect Identifiers in Medical Texts</i> Ibrahim Baroud, Lisa Raithel, Sebastian Möller and Roland Roller	75
<i>Investigating User Perspectives on Differentially Private Text Privatization</i> Stephen Meisenbacher, Alexandra Klymenko, Alexander Karpp and Florian Matthes	86