

The Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025)

Online
31 July - 1 August

ISBN: 979-8-3313-2420-9

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© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

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

<i>From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages</i>	
Artur Kiulian, Anton Polishko, Mykola Khandoga, Yevhen Kostiuk, Guillermo Gabrielli, Łukasz Gałała, Fadi Zaraket, Qusai Abu Obaida, Hrishikesh Garud, Wendy Wing Yee Mak, Dmytro Chaplynskyi, Selma Amor and Grigol Peradze	1
<i>Benchmarking Multimodal Models for Ukrainian Language Understanding Across Academic and Cultural Domains</i>	
Yurii Paniv, Artur Kiulian, Dmytro Chaplynskyi, Mykola Khandoga, Anton Polishko, Tetiana Bas and Guillermo Gabrielli	14
<i>Improving Named Entity Recognition for Low-Resource Languages Using Large Language Models: A Ukrainian Case Study</i>	
Vladyslav Radchenko and Nazarii Drushchak	27
<i>UAlign: LLM Alignment Benchmark for the Ukrainian Language</i>	
Andrian Kravchenko, Yurii Paniv and Nazarii Drushchak	36
<i>Comparing Methods for Multi-Label Classification of Manipulation Techniques in Ukrainian Telegram Content</i>	
Oleh Melnychuk	45
<i>Framing the Language: Fine-Tuning Gemma 3 for Manipulation Detection</i>	
Mykola Khandoga, Yevhen Kostiuk, Anton Polishko, Kostiantyn Kozlov, Yurii Filipchuk and Artur Kiulian	49
<i>Developing a Universal Dependencies Treebank for Ukrainian Parliamentary Speech</i>	
Maria Shvedova, Arsenii Lukashevskyi and Andriy Rysin	55
<i>GBEM-UA: Gender Bias Evaluation and Mitigation for Ukrainian Large Language Models</i>	
Mykhailo Buleshnyi, Maksym Buleshnyi, Marta Sumyk and Nazarii Drushchak	64
<i>A Framework for Large-Scale Parallel Corpus Evaluation: Ensemble Quality Estimation Models Versus Human Assessment</i>	
Dmytro Chaplynskyi and Kyrylo Zakharov	73
<i>Vuyko Mistral: Adapting LLMs for Low-Resource Dialectal Translation</i>	
Roman Kyslyi, Yuliia Maksymiuk and Ihor Pysmennyi	86
<i>Context-Aware Lexical Stress Prediction and Phonemization for Ukrainian TTS Systems</i>	
Anastasiia Senyk, Mykhailo Lukianchuk, Valentyna Robeiko and Yurii Paniv	96
<i>The UNLP 2025 Shared Task on Detecting Social Media Manipulation</i>	
Roman Kyslyi, Nataliia Romanyshyn and Volodymyr Sydorskyi	105
<i>Transforming Causal LLM into MLM Encoder for Detecting Social Media Manipulation in Telegram</i>	
Anton Bazdyrev, Ivan Bashtovyi, Ivan Havlytskyi, Oleksandr Kharytonov and Artur Khodakovskyi	112
<i>On the Path to Make Ukrainian a High-Resource Language</i>	
Mykola Haltiuk and Aleksander Smywiński-Pohl	120
<i>Precision vs. Perturbation: Robustness Analysis of Synonym Attacks in Ukrainian NLP</i>	
Volodymyr Mudryi and Oleksii Ignatenko	131

<i>Gender Swapping as a Data Augmentation Technique: Developing Gender-Balanced Datasets for Ukrainian Language Processing</i>	
Olha Nahurna and Mariana Romanyshyn	147
<i>Introducing OmniGEC: A Silver Multilingual Dataset for Grammatical Error Correction</i>	
Roman Kovalchuk, Mariana Romanyshyn and Petro Ivaniuk	162
<i>Improving Sentiment Analysis for Ukrainian Social Media Code-Switching Data</i>	
Yurii Shynkarov, Veronika Solopova and Vera Schmitt	179
<i>Hidden Persuasion: Detecting Manipulative Narratives on Social Media During the 2022 Russian Invasion of Ukraine</i>	
Kateryna Akhynko, Oleksandr Kosovan and Mykola Trokhymovych	194
<i>Detecting Manipulation in Ukrainian Telegram: A Transformer-Based Approach to Technique Classification and Span Identification</i>	
Md. Abdur Rahman and Md Ashiqur Rahman	203