UNLP 2025

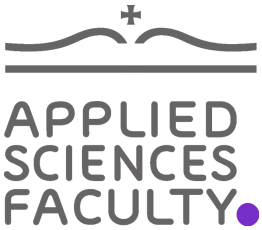
The Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025)

Proceedings of the Workshop

July 31 - August 1, 2025

The UNLP organizers gratefully acknowledge the support from the following sponsors.

UNLP 2025 Partners:



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 979-8-89176-269-5

Welcome to UNLP 2025

We warmly welcome you to the Fourth Ukrainian Natural Language Processing Workshop, held on July 31–August 1, 2025, in conjunction with ACL 2025!

The workshop brings together leading professionals from academia and industry who develop language resources, tools, and NLP solutions for the Ukrainian language. UNLP provides a platform for discussion and sharing of ideas, fosters collaboration between different research groups, and improves the visibility of the Ukrainian research community worldwide.

This year, the workshop received a record 41 submissions, of which 20 were accepted to be presented at the workshop. The paper topics follow the global NLP trends and focus on the customization and application of large language models to a variety of tasks in Ukrainian. Almost half of the papers introduce new large-scale silver datasets for training and fine-grained golden datasets for benchmarking. We were excited to accept three papers in the area of responsible AI, which tackle gender bias and the ethical issues of generative AI. We are immensely grateful to the program committee for their careful and thoughtful reviews of the papers submitted this year!

UNLP 2025 will host two keynote speeches. Sebastian Ruder, Research Scientist at Meta, will discuss the multilingual modeling methods and evaluations the team used for Llama 4 and the current challenges in cross-lingual research, specifically focusing on Ukrainian. Illia Strelnykov, Data Scientist at YouScan, will focus on leveraging user feedback to enhance model performance, addressing such challenges as noise in user data, bias, and conflicting information.

The fourth UNLP will feature the Shared Task on Detecting Social Media Manipulation. This shared task aims to challenge and assess AI capabilities to detect and classify manipulation, laying the groundwork for progress in cybersecurity and the identification of disinformation within the context of Ukraine. The shared task included two tracks: technique classification and span identification. Twenty-two teams submitted their solutions, and five shared task papers were accepted for presentation at the workshop.

To extensively cover the timely topic of manipulation and disinformation, UNLP 2025 will also host a panel discussion on disinformation detection with industry experts from LetsData, Texty.org.ua, Osavul, and OpenMinds.

We express our gratitude to Grammarly for financial and promotional support, Texty.org.ua for providing the dataset for the shared task, UCU's Faculty of Applied Sciences for hosting the UNLP event at the premises of the university, and NaUKMA's Faculty of Computer Sciences for technical support.

We are looking forward to the workshop and anticipate lively discussions on Ukrainian NLP!

Organizers of UNLP 2025,
Mariana Romanyshyn, Olena Nahorna, Oleksii Ignatenko, Andrii Hlybovets

Organizing Committee

Workshop Organizing Committee

Mariana Romanyshyn, Grammarly, Ukraine
Olena Nahorna, Grammarly, Germany
Oleksii Ignatenko, Ukrainian Catholic University, Ukraine
Andrii Hlybovets, National University of Kyiv-Mohyla Academy, Ukraine

Shared Task Organizing Committee

Nataliia Romanyshyn, Ukrainian Catholic University, Texty.org.ua, Ukraine
Roman Kyslyi, Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
Volodymyr Sydorskyi, Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

Program Committee

Program Committee

Bogdan Babych, Heidelberg University, Germany
Anton Bazdyrev, Dun&Bradstreet, Ukraine
Nataliia Cheilytko, Friedrich Schiller University Jena, Germany
Artem Chernodub, ZenDesk, Poland
Nazarii Drushchak, Ukrainian Catholic University, Ukraine
Svitlana Galeshchuk, Université Paris Dauphine, France; West Ukrainian National University, Ukraine
Natalia Grabar, CNRS, Université de Lille, France
Thierry Hamon, Université Paris-Saclay, CNRS, LIMSI & Université Sorbonne, France
Serhii Hamotskyi, Anhalt University of Applied Sciences, Germany
Serhii Havrylov, University of Edinburgh, UK
Oleksiy Syvokon, Lviv Polytechnic National University, Ukraine
Olha Kanishcheva, Friedrich Schiller University Jena, Germany
Natalia Kotsyba, Samsung Research, Poland
Taras Lehinevych, Amazon, Ireland
Andrii Liubonko, EPAM Systems, Ukraine
Iuliia Makogon, Newxel, Ukraine
Oleksandr Marchenko, Taras Shevchenko National University of Kyiv, Ukraine
Oleksii Molchanovskyi, Ukrainian Catholic University, Ukraine
Mark Norris, Grammarly, USA
Kostiantyn Omelianchuk, Grammarly, Germany
Yurii Paniv, Ukrainian Catholic University, Ukraine
Nataliya Polyakovska, SoftServe, USA
Anna Rogers, IT University of Copenhagen, Denmark
Igor Samokhin, Adimen, Ukraine
Tatjana Scheffler, Ruhr-Universität Bochum, Germany
Uta Seewald-Heeg, Anhalt University of Applied Sciences, Germany
Taras Shevchenko, Proxet, Ukraine
Maria Shvedova, National Technical University “Kharkiv Polytechnic Institute”, Ukraine; Friedrich Schiller University Jena, Germany
Oleksandr Skurzhanskyi, Grammarly, Germany
Veronika Solopova, Technische Universität Berlin, Germany
Vasyl Starko, Ukrainian Catholic University, Ukraine
Volodymyr Taranukha, Taras Shevchenko National University of Kyiv, Ukraine
Maksym Tarnavskyi, Shelf, Poland
Oleksii Turuta, Kharkiv National University of Radio Electronics, Ukraine
Taras Ustyianovych, Lviv Polytechnic National University, Ukraine

Invited Talk

Multilingual Modeling and Evaluation in Llama 4 and Beyond

Sebastian Ruder
Meta, Germany

Thursday, July 31, 2025 – Time: **12:00 – 13:00** – Room: **online**

Abstract: In this talk, I will cover some of the multilingual modeling methods and evaluations we used for Llama 4. Looking ahead, I will discuss the current challenges in cross-lingual research, with a focus on Ukrainian specifically.

Invited Talk

Leveraging User Feedback to Improve Your Models

Illia Strelnykov
YouScan, Ukraine

Thursday, July 31, 2025 – Time: **16:00 – 17:00** – Room: **online**

Abstract: While academic research provides a strong foundation for model development, the ultimate goal is to deploy these models in real-world applications, where they interact with actual users. This talk addresses the critical challenge of effectively leveraging user feedback to enhance model performance in practical scenarios. We'll explore ways to incorporate the highly valuable — yet inherently noisy —

user-provided data into model training and fine-tuning pipelines. First, we'll cover methods for collecting user feedback and the challenges involved in processing it, including issues like bias and conflicting information. Then we will examine various solutions for tackling these challenges and how to use refined feedback for model improvement.

Table of Contents

Program

Thursday, July 31, 2025

09:00 - 09:10 *Opening Remarks*

09:10 - 10:30 *Morning Session: Downstream Tasks*

Improving Named Entity Recognition for Low-Resource Languages Using Large Language Models: A Ukrainian Case Study
Vladyslav Radchenko and Nazarii Drushchak

A Framework for Large-Scale Parallel Corpus Evaluation: Ensemble Quality Estimation Models Versus Human Assessment
Dmytro Chaplynskyi and Kyrylo Zakharov

Introducing OmniGEC: A Silver Multilingual Dataset for Grammatical Error Correction
Roman Kovalchuk, Mariana Romanyshyn and Petro Ivaniuk

Improving Sentiment Analysis for Ukrainian Social Media Code-Switching Data
Yurii Shynkarov, Veronika Solopova and Vera Schmitt

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

11:00 - 12:00 *Morning Session: Towards a Ukrainian LLM*

From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages
Artur Kiulian, Anton Polishko, Mykola Khandoga, Yevhen Kostiuk, Guillermo Gabrielli, Łukasz Gagała, Fadi Zaraket, Qusai Abu Obaida, Hrishikesh Garud, Wendy Wing Yee Mak, Dmytro Chaplynskyi, Selma Amor and Grigol Peradze

Benchmarking Multimodal Models for Ukrainian Language Understanding Across Academic and Cultural Domains
Yurii Paniv, Artur Kiulian, Dmytro Chaplynskyi, Mykola Khandoga, Anton Polishko, Tetiana Bas and Guillermo Gabrielli

On the Path to Make Ukrainian a High-Resource Language
Mykola Haltiuk and Aleksander Smywiński-Pohl

12:00 - 13:00 *Keynote: Sebastian Ruder, "Multilingual Modeling and Evaluation in Llama 4 and Beyond"*

13:00 - 14:15 *Lunch*

Thursday, July 31, 2025 (continued)

14:15 - 15:30 *Afternoon Session: Linguistics and NLP*

Developing a Universal Dependencies Treebank for Ukrainian Parliamentary Speech
Maria Shvedova, Arsenii Lukashevskyi and Andriy Rysin

Vuyko Mistral: Adapting LLMs for Low-Resource Dialectal Translation
Roman Kyslyi, Yuliia Maksymiuk and Ihor Pysmennyi

Context-Aware Lexical Stress Prediction and Phonemization for Ukrainian TTS Systems
Anastasiia Senyk, Mykhailo Lukianchuk, Valentyna Robeiko and Yurii Paniv

Precision vs. Perturbation: Robustness Analysis of Synonym Attacks in Ukrainian NLP
Volodymyr Mudryi and Oleksii Ignatenko

15:30 - 16:00 *Afternoon Coffee Break*

16:00 - 17:00 *Keynote: Illia Strelnykov, "Leveraging User Feedback to Improve Your Models"*

17:00 - 18:00 *Afternoon Session: Responsible AI*

UAlign: LLM Alignment Benchmark for the Ukrainian Language
Andrian Kravchenko, Yurii Paniv and Nazarii Drushchak

GBEM-UA: Gender Bias Evaluation and Mitigation for Ukrainian Large Language Models
Mykhailo Buleshnyi, Maksym Buleshnyi, Marta Sumyk and Nazarii Drushchak

Gender Swapping as a Data Augmentation Technique: Developing Gender-Balanced Datasets for Ukrainian Language Processing
Olha Nahurna and Mariana Romanyshyn

17:50 - 18:00 *Closing Words*

Friday, August 1, 2025

09:00 - 10:30 *Morning Session: Downstream Tasks*

The UNLP 2025 Shared Task on Detecting Social Media Manipulation
Roman Kyslyi, Nataliia Romanyshyn and Volodymyr Sydorskyi

Detecting Manipulation in Ukrainian Telegram: A Transformer-Based Approach to Technique Classification and Span Identification
Md. Abdur Rahman and Md Ashiqur Rahman

Hidden Persuasion: Detecting Manipulative Narratives on Social Media During the 2022 Russian Invasion of Ukraine
Kateryna Akhynko, Oleksandr Kosovan and Mykola Trokhymovych

Comparing Methods for Multi-Label Classification of Manipulation Techniques in Ukrainian Telegram Content
Oleh Melnychuk

Framing the Language: Fine-Tuning Gemma 3 for Manipulation Detection
Mykola Khandoga, Yevhen Kostiuk, Anton Polishko, Kostiantyn Kozlov, Yurii Filipchuk and Artur Kiulian

Transforming Causal LLM into MLM Encoder for Detecting Social Media Manipulation in Telegram
Anton Bazdyrev, Ivan Bashtovyi, Ivan Havlytskyi, Oleksandr Kharytonov and Artur Khodakovskyi

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

11:00 - 12:50 *Panel Discussion: "Disinformation Detection from a Business Perspective"*

12:50 - 13:00 *Closing Words*