

UNLP 2026

**The Fifth Ukrainian Natural Language Processing  
Conference (UNLP 2026)**

**Proceedings of the Conference**

May 29-30, 2026

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

**UNLP 2026 Partners:**



©2026 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](mailto:acl@aclweb.org)

ISBN 979-8-89176-359-3

## Welcome to UNLP 2026

We warmly welcome you to the Fifth Ukrainian Natural Language Processing Conference, held on May 29–30, 2026!

The conference 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 conference expanded the list of topics to include more aspects of Ukrainian language processing, such as multimodality, speech, optical character recognition, agentic systems, educational tools, and more. We received a record 52 submissions, of which 22 were accepted to be presented at the conference. 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 datasets for training and benchmarking. We are immensely grateful to the program committee for their careful and thoughtful reviews of the papers submitted this year!

UNLP 2026 will host three keynote speeches. Anna Rogers, Associate Professor at IT University of Copenhagen, will discuss the future of NLP in the age of large language models. Mykola Haltiuk, PhD Student at AGH University of Krakow, will present his work on model-aware tokenizer transfer. Yurii Paniv, PhD Student at Ukrainian Catholic University, will explore approaches to automating research.

The fifth UNLP will feature the Shared Task on Multi-Domain Document Understanding. The shared task challenges AI systems to identify relevant information in collections of domain-specific documents and to generalize across domains. Fifteen teams submitted their solutions, and four shared task papers were accepted for presentation at the conference.

To address the topic of responsible AI, UNLP 2026 will host a panel discussion on ethics of AI usage with Oleksii Molchanovskyi, Chief Innovation Officer at Ukrainian Catholic University, and Olena Andriienko, Chief LegalTech Officer at Publicis Groupe.

We are grateful to the Applied Sciences Faculty of the Ukrainian Catholic University for hosting the conference. We thank MacPaw for its financial support, Preply for labeling the shared task dataset, and Superhuman for providing Grammarly Pro to UNLP authors. Our thanks go to NaUKMA's Faculty of Computer Sciences for technical support. Finally, we thank Kyiv School of Economics, Texty.org.ua, AI House, and all our partners for their promotional and outreach support.

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

Organizers of UNLP 2026,

Roman Kyslyi, Mariana Romanyshyn, Olena Nahorna, Oleksii Ignatenko, and Andrii Hlybovets

# **Organizing Committee**

## **Workshop Organizing Committee**

Roman Kyslyi, Kyiv School of Economics, Ukraine

Mariana Romanyshyn, Superhuman, Ukraine

Olena Nahorna, Preply, Germany

Oleksii Ignatenko, Ukrainian Catholic University, Ukraine

Andrii Hlybovets, National University of Kyiv-Mohyla Academy, Ukraine

## **Shared Task Organizing Committee**

Roman Kyslyi, Kyiv School of Economics, Ukraine

Mariana Romanyshyn, Superhuman, Ukraine

Olena Nahorna, Preply, Germany

Volodymyr Sydorskyi, Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”,  
Ukraine

Nataliia Romanyshyn, Ukrainian Catholic University, Texty.org.ua, Ukraine

# Program Committee

## Program Committee

Andrii Liubonko, EPAM Systems, Ukraine  
Andrii Yaroshevskiy, Respeecher, Ukraine  
Anna Rogers, IT University of Copenhagen, Denmark  
Anton Bazdyrev, Dun&Bradstreet, Ukraine  
Artem Chernodub, Zendesk, Poland  
Bogdan Babych, Heidelberg University, Germany  
Bohdan Didenko, WebSpellChecker LLC, Ukraine; Lviv Polytechnic National University, Ukraine  
Hasan Abu-Rasheed, Goethe University Frankfurt, Germany  
Igor Samokhin, Adimen, Ukraine  
Iuliia Makogon, Newxel, Ukraine  
Kostiantyn Omelianchuk, Superhuman, Germany  
Maksym Tarnavskiy, Shelf, Poland  
Maria Shvedova, National Technical University “Kharkiv Polytechnic Institute”, Ukraine; Friedrich Schiller University Jena, Germany  
Mark Norris, Superhuman, USA  
Mykola Khandoga, AryaXAI, France  
Mykola Sazhok, CyberMova, Ukraine; Institute of Information Technologies and Systems of the NAS of Ukraine  
Natalia Grabar, CNRS, Université de Lille, France  
Natalia Kotsyba, Samsung Research Poland, Poland  
Nataliia Cheilytko, Friedrich Schiller University Jena, Germany  
Nataliya Polyakovska, SoftServe, USA  
Nazarii Drushchak, Ukrainian Catholic University, Ukraine  
Oleksandr Marchenko, Taras Shevchenko National University of Kyiv, Ukraine  
Oleksandr Skurzhanyskiy, Superhuman, Germany  
Oleksii Molchanovskiy, Ukrainian Catholic University, Ukraine  
Oleksii Turuta, Kharkiv National University of Radio Electronics, Ukraine  
Oleksiy Syvokon, Lviv Polytechnic National University, Ukraine  
Olha Kanishcheva, Friedrich Schiller University Jena, Germany  
Serhii Hamotskiy, Anhalt University of Applied Sciences, Germany  
Svitlana Galeshchuk, Université Paris Dauphine, France; West Ukrainian National University, Ukraine  
Taras Lehinevych, Amazon, Ireland  
Taras Shevchenko, Proxet, Ukraine  
Taras Ustyianovych, Lviv Polytechnic National University, Ukraine  
Tatjana Scheffler, Ruhr-Universität Bochum, Germany  
Thierry Hamon, Université Paris-Saclay, CNRS, LIMSIS & Université Sorbonne, France  
Uta Seewald-Heeg, Anhalt University of Applied Sciences, Germany  
Valentyna Robeiko, Taras Shevchenko National University of Kyiv, Ukraine  
Vera Schmitt, Technische Universität Berlin, Germany  
Veronika Solopova, Technische Universität Berlin, Germany  
Volodymyr Kyrlov, OpenAI, USA  
Volodymyr Mudryi, Ukrainian Catholic University, Ukraine  
Volodymyr Taranukha, Taras Shevchenko National University of Kyiv, Ukraine  
Yevhen Kostyuk, Aarhus University, Denmark  
Yevhenii Azarov, Respeecher, Ukraine

Yurii Paniv, Ukrainian Catholic University, Ukraine

## Table of Contents

<i>Improving Domain-Specific Translation from English into Ukrainian with Retrieval-Augmented Generation</i>	
Anton Shpigunov .....	1
<i>UAReviews: A Multi-Task Ukrainian Dataset for Emotion and Intent Classification</i>	
Roman Kyslyi, Ihor Pysmennyi and Denys Mykhailov .....	12
<i>A Two-Axis Framework for Analyzing Ukrainian Dialogues</i>	
Artem Korotenko and Roman Kyslyi .....	24
<i>Entropy of Ukrainian</i>	
Anton Lavreniuk, Mykyta Mudryi and Markiian Chaklosh .....	33
<i>SimIdioms: A Corpus and Benchmark for Ukrainian Idiom Translation</i>	
Yaryna Petruniv, Iuliia Makogon and Roman Kyslyi .....	41
<i>UkrSL: Towards a Ukrainian Continuous Sign Language Dataset</i>	
Oleksandr Sobetskyi, Maryna Kosse, Roman Kyslyi and Angelina Savchenko .....	53
<i>Digitizing Old Ukrainian Texts: A Prompt-Based OCR Pipeline and Evaluation Dataset</i>	
Dmytro Chaplynskyi and Hanna Dydyk-Meush .....	58
<i>Semantic Fidelity Versus Literary Quality: A Construct Validity Study of Neural Machine Translation Metrics</i>	
Dmytro Chaplynskyi, Ivan Kulynych, Maria Shvedova and Lesia Ivashkevych .....	67
<i>Graph-Based Detection of Disinformation Narrative Diffusion between Russian and Ukrainian Telegram Channels</i>	
Yuliia Vistak, Viktoriia Makovska, Vera Schmitt and Veronika Solopova .....	80
<i>Professional Translators Versus Quality Estimation Models: Reliability and Agreement in English-Ukrainian Translation Evaluation</i>	
Dmytro Chaplynskyi, Kyrylo Zakharov and Lesia Ivashkevych .....	97
<i>Belief Propagation in LLM World Models: Measuring Strategic Information Bias with Prediction Markets</i>	
Mykola Khandoga, Yevhen Kostiuk, Anton Polishko, Yurii Filipchuk, Kostiantyn Kozlov, Dmytro Zamriy and Artur Kiulian .....	108
<i>Toward a Gold-Standard Benchmark for Evaluating Ukrainian Language Proficiency in LLMs</i>	
Svitlana Galeshchuk, Yuliia Maksymiuk, Yuliia Chernobrov, Nina Stankevych, Oleksandra Antoniv, Nataliia Faryna and Oksana Popkova .....	121
<i>How Far Can Prompting Go for Minimal-Edit Ukrainian Grammatical Error Correction?</i>	
Kateryna Karpo and Artem Chernodub .....	136
<i>Data-Efficient Adaptation of Multilingual LLMs to Ukrainian</i>	
Yurii Paniv, Bohdan Didenko, Mykola Haltiuk, Vladyslav Humennyi, Andrian Kravchenko, Roman Kyslyi, Viktoriia Makovska, Artem Orlovskyi, Bohdan Ruban, Maksym-Yurii Rudko, Anastasiia Senyk, Nazarii Drushchak, Dmytro Chaplynskyi and Mariana Romanyshyn .....	155
<i>Dictionary-Based Speculative Decoding for Non-Latin-Script Languages</i>	
Oleksiy Syvokon .....	169

<i>Scaling ASR for Hutsul Dialect: Multi-Speaker Data Collection, Enhanced Transcription and Cross-Speaker Evaluation</i>	
Artem Orlovskyi, Zakhar Guzii, Bohdan Onyshchenko, Roman Kyslyi and Pavlo Khomenko .	184
<i>Mining Native Ukrainian Paraphrases: A Multi-Source Comparison</i>	
Vladyslav Fesenko, Hanna Dydyk-Meush and Volodymyr Mudryi . . . . .	199
<i>Automated CEFR-Level Assignment for Ukrainian Texts</i>	
Olha Kanishcheva and Mikhail Kopotev . . . . .	209
<i>An End-to-End Ukrainian RAG for Local Deployment. Optimized Hybrid Search and Lightweight Generation</i>	
Mykola Trokhymovych, Yana Oliinyk and Nazarii Nyzhnyk . . . . .	223
<i>Qwen Goes Brrr: Off-the-Shelf RAG for Ukrainian Multi-Domain Document Understanding</i>	
Anton Bazdyrev, Oleksandr Kharytonov, Artur Khodakovskiy, Ivan Havlytskyi and Ivan Bashtovyi . . . . .	230
<i>RAG Pipeline Strategies for Ukrainian Multi-Domain Document Understanding Task</i>	
Mykola Nosenko and Pavlo Kilko . . . . .	240
<i>The UNLP 2026 Shared Task on Multi-Domain Document Understanding</i>	
Volodymyr Sydorskyi, Nataliia Romanyshyn, Roman Kyslyi and Olena Nahorna . . . . .	249

# Program

**Friday, May 29, 2026**

10:00 - 10:15     *Opening Remarks*

10:15 - 11:15     *Morning Session: Translation & Evaluation*

*SimIdioms: A Corpus and Benchmark for Ukrainian Idiom Translation*

Yaryna Petruniv, Iuliia Makogon and Roman Kyslyi

*Semantic Fidelity Versus Literary Quality: A Construct Validity Study of Neural Machine Translation Metrics*

Dmytro Chaplynskyi, Ivan Kulynych, Maria Shvedova and Lesia Ivashkevych

*Professional Translators Versus Quality Estimation Models: Reliability and Agreement in English-Ukrainian Translation Evaluation*

Dmytro Chaplynskyi, Kyrylo Zakharov and Lesia Ivashkevych

11:15 - 11:45     *Morning Coffee Break*

11:45 - 13:05     *Morning Session: Core NLP & Linguistic Resources*

*Data-Efficient Adaptation of Multilingual LLMs to Ukrainian*

Yurii Paniv, Bohdan Didenko, Mykola Haliuk, Vladyslav Humennyi, Andrian Kravchenko, Roman Kyslyi, Viktoriia Makovska, Artem Orlovskyi, Bohdan Ruban, Maksym-Yurii Rudko, Anastasiia Senyk, Nazarii Drushchak, Dmytro Chaplynskyi and Mariana Romanyshyn

*Dictionary-Based Speculative Decoding for Non-Latin-Script Languages*

Oleksiy Syvokon

*How Far Can Prompting Go for Minimal-Edit Ukrainian Grammatical Error Correction?*

Kateryna Karpo and Artem Chernodub

*Entropy of Ukrainian*

Anton Lavreniuk, Mykyta Mudryi and Markiiian Chaklosh

13:05 - 14:15     *Lunch*

14:15 - 15:15     *Keynote: Anna Rogers, “What’s Next for NLP after LLMs?”*

**Friday, May 29, 2026 (continued)**

15:15 - 16:15 *Afternoon Session: Language Proficiency & Paraphrasing*

*Toward a Gold-Standard Benchmark for Evaluating Ukrainian Language Proficiency in LLMs*

Svitlana Galeshchuk, Yuliia Maksymiuk, Yuliia Chernobrov, Nina Stankevych, Oleksandra Antoniv, Nataliia Faryna and Oksana Popkova

*Automated CEFR-Level Assignment for Ukrainian Texts*

Olha Kanishcheva and Mikhail Kopotev

*Mining Native Ukrainian Paraphrases: A Multi-Source Comparison*

Vladyslav Fesenko, Hanna Dydyk-Meush and Volodymyr Mudryi

16:15 - 16:45 *Afternoon Coffee Break*

16:45 - 18:00 *Keynote: Mykola Haliuk, “Model-Aware Tokenizer Transfer”*

18:00 - 18:10 *End of Day 1*

**Saturday, May 30, 2026**

10:00 - 10:10 *Day 2 Welcome*

10:10 - 11:10 *Morning Session: Speech, Multimodality & OCR*

*Scaling ASR for Hutsul Dialect: Multi-Speaker Data Collection, Enhanced Transcription and Cross-Speaker Evaluation*

Artem Orlovskiy, Zakhar Guzii, Bohdan Onyshchenko, Roman Kyslyi and Pavlo Khomenko

*UkrSL: Towards a Ukrainian Continuous Sign Language Dataset*

Oleksandr Sobetskyi, Maryna Kosse, Roman Kyslyi and Angelina Savchenko

*Digitizing Old Ukrainian Texts: A Prompt-Based OCR Pipeline and Evaluation Dataset*

Dmytro Chaplynskyi and Hanna Dydyk-Meush

11:10 - 11:40 *Morning Coffee Break*

11:40 - 13:00 *Morning Session: Applications, Systems & Tools*

*Improving Domain-Specific Translation from English into Ukrainian with Retrieval-Augmented Generation*

Anton Shpigunov

*UARreviews: A Multi-Task Ukrainian Dataset for Emotion and Intent Classification*

Roman Kyslyi, Ihor Pysmennyi and Denys Mykhailov

*Graph-Based Detection of Disinformation Narrative Diffusion between Russian and Ukrainian Telegram Channels*

Yuliia Vistak, Viktoriia Makovska, Vera Schmitt and Veronika Solopova

*A Two-Axis Framework for Analyzing Ukrainian Dialogues*

Artem Korotenko and Roman Kyslyi

13:00 - 14:15 *Lunch*

14:15 - 14:35 *Afternoon Session: Applications, Systems & Tools (continued)*

**Saturday, May 30, 2026 (continued)**

*Belief Propagation in LLM World Models: Measuring Strategic Information Bias with Prediction Markets*

Mykola Khandoga, Yevhen Kostiuk, Anton Polishko, Yurii Filipchuk, Kostiantyn Kozlov, Dmytro Zamriy and Artur Kiulian

14:35 - 15:35 *Keynote: Yurii Paniv, “On Automating Research”*

15:35 - 16:55 *Afternoon Session: Shared Task on Multi-Domain Document Understanding*

*The UNLP 2026 Shared Task on Multi-Domain Document Understanding*

Volodymyr Sydorskyi, Nataliia Romanyshyn, Roman Kyslyi and Olena Nahorna

*RAG Pipeline Strategies for Ukrainian Multi-Domain Document Understanding Task*

Mykola Nosenko and Pavlo Kilko

*An End-to-End Ukrainian RAG for Local Deployment. Optimized Hybrid Search and Lightweight Generation*

Mykola Trokhymovych, Yana Oliinyk and Nazarii Nyzhnyk

*Qwen Goes Brrr: Off-the-Shelf RAG for Ukrainian Multi-Domain Document Understanding*

Anton Bazdyrev, Oleksandr Kharytonov, Artur Khodakovskiy, Ivan Havlytskyi and Ivan Bashtovyi

17:00 - 18:00 *Panel Discussion: Oleksii Molchanovskiy, Olena Andriienko, and Roman Kyslyi, “Ethics of AI Usage”*

18:00 - 18:15 *Closing Words*