

UncertaiNLP 2025

**Workshop on Uncertainty-Aware NLP (UncertaiNLP 2025)**

**Proceedings of the Workshop**

November 9, 2025

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



©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](mailto:acl@aclweb.org)

ISBN 979-8-89176-349-4

## Introduction

Human languages are inherently ambiguous and understanding language input is subject to interpretation and complex contextual dependencies. Nevertheless, the main body of research in NLP is still based on the assumption that ambiguities and other types of underspecification can and have to be resolved. This second edition of the Uncertainty-Aware NLP workshop (UncertaiNLP 2025) provides a platform for research that embraces variability in human language and aims to represent and evaluate the uncertainty that arises both from language itself and from the modeling tools we use to process it.

Uncertainty arises when multiple outcomes are possible and variability cannot be fully explained by context. In NLP, it reflects both epistemic and aleatoric factors, stemming from linguistic ambiguity, individual variation, domain shifts, and modeling choices. These challenges are especially acute in low-resource settings, where over- and under-fitting risks increase. Consequently, uncertainty-aware NLP research spans model design, data collection, inference, and evaluation, with growing importance in mission-critical applications requiring reliable confidence estimation. The success of the UncertaiNLP workshop’s inaugural edition at EACL 2024 and the expanding community around related themes, underscores the growing research interest in this area and the need for continued exchange and collaboration. Building on that momentum, the second edition of UncertaiNLP nearly doubled in size, reflecting the field’s rapid growth and the community’s increasing recognition of the importance of uncertainty-aware methods in NLP.

This volume contains the proceedings of the second edition of the UncertaiNLP workshop hosted on November 9th, 2025, co-located with the 2025 Conference on Empirical Methods in Natural Language Processing in the Suzhou International Expo Centre in Suzhou, Jiangsu Province, China. We invited paper submissions on a wide variety of topics, including representing, documenting or modeling uncertainty, parameter estimation, probabilistic inference, decision making, evaluation and calibration, and hallucinations and uncertainty-driven mitigation. We received a total of 50 submissions, of which we accepted 17 long and 10 short papers, amounting to an acceptance rate of 54%

We are grateful to our invited keynote speakers: Gal Yona (Google Research, IL), Maxim Panov (MB-ZUAI, UAE), Parisa Kordjamshidi (Michigan State University, USA), Eyke Hüllermeier (LMU Munich, DE). We would also like to thank the EU’s Horizon Europe research and innovation program for support through the Unified Transcription and Translation (UTTER, agreement No. 101070631) and the Foundation for Empirical Multimodality Research (FOUNDATIONS, agreement No. 101122047) projects. This workshop is also partially supported by an unrestricted gift from Google (Google research scholar award).

The UncertaiNLP organizers, Wilker Aziz, Jonathan Berant, Bryan Eikema, Marie-Catherine de Marneffe, Barbara Plank, Artem Shelmanov, Swabha Swayamdipta, Jörg Tiedemann, Raúl Vázquez, Chrysoula Zerva.

# Program Committee

## Program Chairs

Wilker Aziz, University of Amsterdam, Netherlands  
Jonathan Berant, Tel Aviv University and Google DeepMind, Israel  
Bryan Eikema, University of Amsterdam, Netherlands  
Marie-Catherine de Marneffe, UCLouvain and FNRS, Belgium  
Barbara Plank, LMU München and IT University of Copenhagen, Germany and Denmark  
Artem Shelmanov, Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates  
Swabha Swayamdipta, University of Southern California, Viterbi School of Engineering, United States  
Jörg Tiedemann, University of Helsinki, Finland  
Raúl Vázquez, University of Helsinki, Finland  
Chrysoula Zerva, Instituto de Telecomunicações, Portugal

## Program Committee

Luigi Acerbi, University of Helsinki  
Roei Aharoni, Google Research  
Alessandro Antonucci, IDSIA  
Henri Aïdasso, Université du Québec  
Samuel Barry, Mistral AI  
Nitay Calderon, Technion  
Juan Cardenas-Cartagena, University of Groningen  
Arie Cattan, Bar Ilan University  
Julius Cheng, University of Cambridge  
Ye-eun Cho, Sungkyunkwan University  
Caio Corro, INSA Rennes  
Nico Daheim, Technische Universität Darmstadt  
Sarkar Snigdha Sarathi Das, Pennsylvania State University  
Vivek Datla, Capital One  
Bonaventure F. P. Dossou, McGill University  
Adam Faulkner, Capital One  
Pedro Lobato Ferreira, University of Amsterdam  
Antske Fokkens, VU University Amsterdam  
Jes Frellsen, Technical University of Denmark  
Thomas L. Griffiths, Princeton University  
Georg Groh, Technical University Munich  
Christian Hardmeier, IT University Copenhagen  
Zhiqi Huang, CapitalOne  
Evgenia Ilia, University of Amsterdam  
Yuu Jinnai, CyberAgent, Inc.  
Mucheol Kim, Chung-Ang University  
Deepak Kumar, Infrd  
Lucie Kunitomo-Jacquin, AIST  
Haau-Sing Li, Technische Universität Darmstadt  
Edison Marrese-Taylor, University of Tokyo  
Yan Meng, University of Amsterdam

Timothee Mickus, University of Helsinki  
Tatiana Passali, Aristotle University of Thessaloniki  
Laura Perez-Beltrachini, University of Edinburgh  
Yuval Pinter, Ben-Gurion University of the Negev  
Timothy Pistotti, University of Auckland  
Alberto Purpura, CapitalOne  
Julian Rodemann, LMU Munich  
Rico Sennrich, University of Zürich  
Arnab Sharma, Paderborn University  
Anthony Sicilia, Northeastern University  
Edwin Simpson, University of Bristol  
Maciej Skorski, University of Warsaw  
Sharmin Sultana, University of Massachusetts at Lowell  
Aarne Talman, University of Helsinki  
Sergey Troshin, University of Amsterdam  
Grigorios Tsoumakas, Aristotle University of Thessaloniki  
Dennis Ulmer, University of Amsterdam  
Teemu Vahtola, University of Helsinki  
Matias Valdenegro-Toro, University of Groningen  
Sami Virpioja, University of Helsinki  
Daniel Vollmers, Paderborn University  
Ryan Wails, Georgetown University  
Li Wang, CapitalOne  
Di Wu, University of Amsterdam  
Yusen Zhang, Pennsylvania State University

### **Publication Chairs**

Raúl Vázquez  
Bryan Eikema

### **Invited Speakers**

Parisa Kordjamshidi, Michigan State University, USA  
Gal Yona, Google Research, USA  
Maxim Panov, MBZUAI, UAE  
Eyke Hüllermeier, LMU München, Germany

## Keynote Talk

# TBA

**Maxim Panov**

MBZUAI, UAE

**2025-09-11 09:10 – Room: Room A207**

**Abstract:** TBA

**Bio:** Maxim Panov is an Assistant Professor at MBZUAI, UAE. Before joining MBZUAI, Panov worked as a research scientist at DATADVANCE Company, where he participated in developing a library of data analysis methods for engineering applications. This library, pSeven, is now used by many companies worldwide, including Airbus, Porsche, Mitsubishi, Toyota, and Limagrain. From 2018, Panov has been an assistant professor at Skolkovo Institute of Science and Technology, Moscow, where he led a statistical machine learning group. Since 2022, he has led an AI theory and algorithms group at the Technology Innovation Institute, Abu Dhabi, UAE. His research interests lie in uncertainty quantification for machine learning model predictions and Bayesian approaches in machine learning. Maxim is leading a research team dedicated to exploring the theoretical foundations of uncertainty quantification and its practical applications. Maxim is also co-leading the development of the LM-Polygraph framework for uncertainty quantification for LLMs. Maxim was a local chair for the ICDM 2024 conference and a recipient of the Best Paper Runner-up Award at the Uncertainty in Artificial Intelligence 2023 conference.

# Keynote Talk

## Reasoning under Uncertainty with Large Multimodal Language Models

**Parisa Kordjamshidi**

Michigan State University, USA

**2025-09-11 13:15** – Room: **Room A207**

**Abstract:** Uncertainty in intelligent models has multiple facets. One aspect concerns a model’s own uncertainty or confidence in its generated outputs. Another pertains to factual knowledge about uncertainty within specific concepts. For example, statements such as “10–20% of lifelong smokers will develop lung cancer” express factual uncertainty derived from statistical data analyses and represented in text. A key research question is whether language models can form and convey such factual uncertainties—integrating information, drawing on their internal knowledge, and aligning this with their confidence when expressing opinions. While addressing this question is highly challenging, I will present our research that explores related directions and the following research question: 1) How do language models understand uncertainty expressions in natural language and perform probabilistic inference over them? 2) How can models be trained to follow the principles of probabilistic reasoning when handling uncertainty in text? 3) How can today’s large models reason over uncertain text? specifically focusing on mapping language into formal probabilistic logic programs?, and finally, in the context of grounding natural language in the visual modality, 4) How can uncertainty in perception be explicitly represented in reasoning? specifically focusing on mappings to differentiable probabilistic programs.

**Bio:** Parisa Kordjamshidi is an Associate Professor of Computer Science and Engineering at Michigan State University. Her research focuses on Natural Language Processing, multimodal reasoning across vision and language, and neuro-symbolic learning. She received her Ph.D. from KU Leuven and conducted postdoctoral research at the University of Illinois Urbana-Champaign. She is a recipient of the NSF CAREER, Amazon Faculty Research, and Fulbright Scholar Awards, and her research team received the NAACL 2025 Outstanding Research Paper Award. Dr. Kordjamshidi serves as Associate Editor of JAIR, Co-editor in chief of ARR (2026), Action Editor for TACL and has held roles in organization committee of major conferences including ACL, NAACL, EACL, EMNLP, ECML-PKDD, and AAAI. Currently, she is a visiting Associate Professor at UCLA spending a part of her sabbatical.



# Keynote Talk

## Beyond Factuality: Improving Trust and Reliability of Large Language Models

Gal Yona

Google Research, USA

2025-09-11 14:45 – Room: Room A207

**Abstract:** Factuality is a cornerstone for trustworthy LLMs, yet despite impressive progress, frontier LLMs still make many confident errors when faced with questions beyond their knowledge boundaries. In this talk I'll present Faithful Response Uncertainty, a different desiderata that shifts the focus away from measuring the number of incorrect statements and towards measuring the alignment between the model's expressed certainty (decisiveness") and intrinsic certainty (confidence"). I'll conclude with a discussion of open problems and possible next steps at the intersection of factuality and uncertainty in frontier LLMs.

**Bio:** Gal Yona is a Research Scientist at Google Research, Tel Aviv, where she is working on improving factuality in large language models, with an emphasis on robustness and uncertainty. Before joining Google, Gal completed her PhD in Computer Science at the Weizmann Institute of Science, developing definitions and algorithms for preventing discrimination in machine learning models. Gal received numerous award during her PhD, including the Google PhD Fellowship in Machine Learning (2021).

# Keynote Talk

## Challenges in Uncertainty Quantification for Large Language Models

**Eyke Hüllermeier**

LMU Munich, Germany

**2025-09-11 16:00** – Room: **Room A207**

**Abstract:** Uncertainty quantification is important in the context of large language models (LLMs) because the outputs produced by these models are often incorrect. However, due to the complexity of language and the numerous sources of uncertainty in textual data, quantifying uncertainty in LLMs is challenging. Indeed, simply transferring existing approaches to uncertainty quantification developed for standard machine learning problems, such as classification and regression, is neither straightforward nor appropriate. This is particularly pertinent to the definition of aleatoric and epistemic uncertainty, and how they are distinguished based on the notion of reducibility. This talk will discuss the challenges of uncertainty quantification for LLMs, propose potential solutions and highlight promising avenues for future research in this emerging field.

**Bio:** Eyke Hüllermeier is a full professor at the Institute of Informatics at LMU Munich, Germany, where he holds the Chair of Artificial Intelligence and Machine Learning. He studied mathematics and business computing, received his PhD in Computer Science from Paderborn University in 1997, and a Habilitation degree in 2002. Before joining LMU, he held professorships at several other German universities (Dortmund, Magdeburg, Marburg, Paderborn) and spent two years as a Marie Curie fellow at the IRIT in Toulouse (France). His research interests are centered around methods and theoretical foundations of artificial intelligence, with a particular focus on machine learning, preference modeling, and reasoning under uncertainty. He has published more than 400 articles on related topics in top-tier journals and major international conferences, and several of his contributions have been recognized with scientific awards. Professor Hüllermeier is Editor-in-Chief of Data Mining and Knowledge Discovery, Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), and serves on the editorial boards of several other AI and machine learning journals. He is currently also the president of EuADS, the European Association for Data Science.

## Table of Contents

<i>Uncertainty-driven Partial Diacritization for Arabic Text</i> Humaid Ali Alblooshi, Artem Shelmanov and Hanan Aldarmaki .....	1
<i>Phases of Uncertainty: Confidence–Calibration Dynamics in Language Model Training</i> Aneesh Durai .....	11
<i>Beyond Human Judgment: A Bayesian Evaluation of LLMs’ Moral Values Understanding</i> Maciej Skorski and Alina Landowska .....	17
<i>Do Large Language Models Know When Not to Answer in Medical QA?</i> Sravanthi Machcha, Sushrita Yerra, Sharmin Sultana, hong yu and Zonghai Yao .....	27
<i>The Geometry of Creative Variability: How Credal Sets Expose Calibration Gaps in Language Models</i> Esteban Garces Arias, Julian Rodemann and Christian Heumann .....	36
<i>Certain but not Probable? Differentiating Certainty from Probability in LLM Token Outputs for Probabilistic Scenarios</i> Autumn Toney and Ryan Wails .....	51
<i>The Benefits of Being Uncertain: Perplexity as a Signal for Naturalness in Multilingual Machine Translation</i> Timothy Pistotti, Michael J. Witbrock, Dr Padriac Amato Tahua O’Leary and Jason Brown ...	61
<i>Asking a Language Model for Diverse Responses</i> Sergey Troshin, Irina Saparina, Antske Fokkens and Vlad Niculae .....	66
<i>Learning to vary: Teaching LMs to reproduce human linguistic variability in next-word prediction</i> Tobias Groot, Salo Lacunes and Evgenia Ilia .....	73
<i>HALLUCINOGEN: Benchmarking Hallucination in Implicit Reasoning within Large Vision Language Models</i> Ashish Seth, Dinesh Manocha and Chirag Agarwal .....	89
<i>Uncertainty in Semantic Language Modeling with PIXELS</i> Stefania Radu, Marco Zullo and Matias Valdenegro-Toro .....	103
<i>Confidence Calibration in Large Language Model-Based Entity Matching</i> Iris Kamsteeg, Juan Cardenas-Cartagena, Floris van Beers, Tsegaye Misikir Tashu and Matias Valdenegro-Toro .....	120
<i>Consensus or Conflict? Fine-Grained Evaluation of Conflicting Answers in Question-Answering</i> Eviatar Nachshoni, Arie Cattan, Shmuel Amar, Ori Shapira and Ido Dagan .....	138
<i>Demystify Verbosity Compensation Behavior of Large Language Models</i> Yusen Zhang, Sarkar Snigdha Sarathi Das and Rui Zhang .....	160
<i>On the Role of Unobserved Sequences on Sample-based Uncertainty Quantification for LLMs</i> Lucie Kunitomo-Jacquín, Edison Marrese-Taylor and Ken Fukuda .....	179
<i>Confidence-Based Response Abstinence: Improving LLM Trustworthiness via Activation-Based Uncertainty Estimation</i> Zhiqi Huang, Vivek Datla, Chenyang Zhu, Alfy Samuel, Daben Liu, Anoop Kumar and Ritesh Soni .....	184

<i>Amortized Bayesian Meta-Learning for Low-Rank Adaptation of Large Language Models</i> Liyi Zhang, Jake C. Snell and Thomas L. Griffiths .....	194
<i>Towards Trustworthy Summarization of Cardiovascular Articles: A Factuality-and-Uncertainty-Aware Biomedical LLM Approach</i> Eleni Partalidou, Tatiana Passali, Chrysoula Zerva, Grigorios Tsoumakas and Sophia Ananiadou	200
<i>Causal Understanding by LLMs: The Role of Uncertainty</i> Oscar William Lithgow-Serrano, Vani Kanjirang and Alessandro Antonucci .....	208
<i>It Depends: Resolving Referential Ambiguity in Minimal Contexts with Commonsense Knowledge</i> Lukas Ellinger and Georg Groh .....	229
<i>Read Your Own Mind: Reasoning Helps Surface Self-Confidence Signals in LLMs</i> Jakub Podolak and Rajeev Verma .....	247
<i>Calibrating Language Models for Neural Ranking under Noisy Supervision with Relaxed Labels</i> Arnab Sharma, Daniel Vollmers and Axel-Cyrille Ngonga Ngomo .....	259
<i>ERGO: Entropy-guided Resetting for Generation Optimization in Multi-turn Language Models</i> Haziq Mohammad Khalid, Athikash Jeyaganthan, Timothy Do, Yicheng Fu, Vasu Sharma, Sean O'Brien and Kevin Zhu .....	273
<i>Towards Open-Ended Discovery for Low-Resource NLP</i> Bonaventure F. P. Dossou and Henri Aidasso .....	287
<i>Can Vision-Language Models Infer Speaker's Ignorance? The Role of Visual and Linguistic Cues</i> Ye-eun Cho and Yunho Maeng .....	298
<i>DeLTa: A Decoding Strategy based on Logit Trajectory Prediction Improves Factuality and Reasoning Ability</i> Yunzhen He, Yusuke Takase and Hidetoshi Shimodaira .....	309
<i>Investigating Factuality in Long-Form Text Generation: The Roles of Self-Known and Self-Unknown</i> Lifu Tu, Rui Meng, Shafiq Joty, Yingbo Zhou and Semih Yavuz .....	322

# Program

## Sunday, November 9, 2025

- 09:00 - 09:10     *Opening Remarks*
- 09:10 - 09:55     *Keynote Talk 1: Maxim Panov*
- 09:55 - 10:30     *Poster lightning round 1 (in-person)*
- 10:30 - 11:00     *Coffee Break*
- 11:00 - 12:15     *In-Person Poster Session*
- 12:15 - 13:15     *Lunch Break*
- 13:15 - 14:00     *Keynote Talk 2: Parisa Kordjamshidi*
- 14:00 - 14:45     *Poster lightning round 2 (virtual)*
- 14:45 - 15:30     *Keynote Talk 3: Gal Yona*
- 15:30 - 16:00     *Coffee Break*
- 16:00 - 16:45     *Keynote Talk 4: Eyke Hüllermeier*
- 16:45 - 17:00     *Closing Remarks*

### Detailed view of lightning rounds

- 09:30 - 10:20     *Poster lightning round 1 (in-person)*

*Uncertainty-driven Partial Diacritization for Arabic Text*

Humaid Ali Alblooshi, Artem Shelmanov and Hanan Aldarmaki

*The Geometry of Creative Variability: How Credal Sets Expose Calibration Gaps in Language Models*

Esteban Garces Arias, Julian Rodemann and Christian Heumann

Sunday, November 9, 2025 (continued)

*The Benefits of Being Uncertain: Perplexity as a Signal for Naturalness in Multilingual Machine Translation*

Timothy Pistotti, Michael J. Witbrock, Dr Padriac Amato Tahua O’Leary and Jason Brown

*Asking a Language Model for Diverse Responses*

Sergey Troshin, Irina Saparina, Antske Fokkens and Vlad Niculae

*Learning to vary: Teaching LMs to reproduce human linguistic variability in next-word prediction*

Tobias Groot, Salo Lacunes and Evgenia Ilia

*Consensus or Conflict? Fine-Grained Evaluation of Conflicting Answers in Question-Answering*

Eviatar Nachshoni, Arie Cattan, Shmuel Amar, Ori Shapira and Ido Dagan

*Confidence-Based Response Abstinence: Improving LLM Trustworthiness via Activation-Based Uncertainty Estimation*

Zhiqi Huang, Vivek Datla, Chenyang Zhu, Alfy Samuel, Daben Liu, Anoop Kumar and Ritesh Soni

*Causal Understanding by LLMs: The Role of Uncertainty*

Oscar William Lithgow-Serrano, Vani Kanjirangat and Alessandro Antonucci

*It Depends: Resolving Referential Ambiguity in Minimal Contexts with Commonsense Knowledge*

Lukas Ellinger and Georg Groh

*Read Your Own Mind: Reasoning Helps Surface Self-Confidence Signals in LLMs*

Jakub Podolak and Rajeev Verma

*Calibrating Language Models for Neural Ranking under Noisy Supervision with Relaxed Labels*

Arnab Sharma, Daniel Vollmers and Axel-Cyrille Ngonga Ngomo

*ERGO: Entropy-guided Resetting for Generation Optimization in Multi-turn Language Models*

Haziq Mohammad Khalid, Athikash Jeyaganthan, Timothy Do, Yicheng Fu, Vasu Sharma, Sean O’Brien and Kevin Zhu

*Can Vision-Language Models Infer Speaker’s Ignorance? The Role of Visual and Linguistic Cues*

Ye-eun Cho and Yunho Maeng

**Sunday, November 9, 2025 (continued)**

*DeLTa: A Decoding Strategy based on Logit Trajectory Prediction Improves Factuality and Reasoning Ability*

Yunzhen He, Yusuke Takase and Hidetoshi Shimodaira

14:00 - 14:45 *Poster lightning round 2 (virtual)*

*Phases of Uncertainty: Confidence–Calibration Dynamics in Language Model Training*

Aneesh Durai

*Beyond Human Judgment: A Bayesian Evaluation of LLMs' Moral Values Understanding*

Maciej Skorski and Alina Landowska

*Do Large Language Models Know When Not to Answer in Medical QA?*

Sravanthi Machcha, Sushrita Yerra, Sharmin Sultana, hong yu and Zonghai Yao

*Certain but not Probable? Differentiating Certainty from Probability in LLM Token Outputs for Probabilistic Scenarios*

Autumn Toney and Ryan Wails

*HALLUCINOGEN: Benchmarking Hallucination in Implicit Reasoning within Large Vision Language Models*

Ashish Seth, Dinesh Manocha and Chirag Agarwal

*Uncertainty in Semantic Language Modeling with PIXELS*

Stefania Radu, Marco Zullo and Matias Valdenegro-Toro

*Confidence Calibration in Large Language Model-Based Entity Matching*

Iris Kamsteeg, Juan Cardenas-Cartagena, Floris van Beers, Tsegaye Misikir Tashu and Matias Valdenegro-Toro

*Demystify Verbosity Compensation Behavior of Large Language Models*

Yusen Zhang, Sarkar Snigdha Sarathi Das and Rui Zhang

*On the Role of Unobserved Sequences on Sample-based Uncertainty Quantification for LLMs*

Lucie Kunitomo-Jacquín, Edison Marrese-Taylor and Ken Fukuda

**Sunday, November 9, 2025 (continued)**

*Amortized Bayesian Meta-Learning for Low-Rank Adaptation of Large Language Models*

Liyi Zhang, Jake C. Snell and Thomas L. Griffiths

*Towards Trustworthy Summarization of Cardiovascular Articles: A Factuality- and-Uncertainty-Aware Biomedical LLM Approach*

Eleni Partalidou, Tatiana Passali, Chrysoula Zerva, Grigorios Tsoumakos and Sophia Ananiadou

*Towards Open-Ended Discovery for Low-Resource NLP*

Bonaventure F. P. Dossou and Henri Aidasso

*Investigating Factuality in Long-Form Text Generation: The Roles of Self-Known and Self-Unknown*

Lifu Tu, Rui Meng, Shafiq Joty, Yingbo Zhou and Semih Yavuz