CHOMPS 2025

Workshop on Confabulation, Hallucinations and Overgeneration in Multilingual and Practical Settings (CHOMPS 2025)

Proceedings of the Workshop

December 23, 2025

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 979-8-89176-308-1

Introduction

Large Language Models (LLMs) have rapidly become integral to applications far beyond core NLP research. Yet their well-known tendency to produce fluent, confident falsehoods remains a major obstacle to safe and equitable deployment. These behaviors are particularly harmful in precision-critical settings such as medicine, law, biotechnology, and education, where accuracy is non-negotiable, and in multilingual contexts where benchmarks, resources, and robust mitigation strategies lag behind high-resource languages. CHOMPS 2025 was created in response to this growing need: to bring together researchers investigating why LLMs "make things up," how we can detect such failures, and what it takes to build models that are measurably more trustworthy across languages, domains, and contexts.

Hallucination, confabulations and overgenerations arise when models produce outputs that are unsupported, unverifiable, or simply fabricated. Their causes span data biases, training dynamics, decoding strategies, and cross-lingual transfer challenges; factors that lead models to generate text that may sound plausible yet be misleading and harmful in practice. Recent shared tasks such as SHROOM and Mu-SHROOM have highlighted just how difficult it remains to detect such errors reliably, especially at scale and in multilingual settings. As LLMs continue to move into high-stakes workflows, understanding the sources, manifestations, and mitigation of hallucinations has become essential for responsible AI development. CHOMPS 2025 aims to foreground this conversation by connecting empirical research on hallucination detection and model behavior with the perspective of practitioners and domain experts who encounter these failures in real-world environments.

This volume contains the proceedings of the inaugural CHOMPS: Workshop on Hallucinations, Confabulations, and Overgeneration in Real-World and Multilingual Settings, held in 2025 and co-located with the International Joint Conference on Natural Language Processing & Asia-Pacific Chapter of the Association for Computational Linguistics (AACL-IJCNLP 2025) in Mumbai, India. We invited submissions on a wide range of topics, including metrics and benchmarks for detecting hallucinations; mitigation techniques at training and inference time; analyses of confabulation in multilingual and multimodal models; and domain-specific case studies from healthcare, law, education, and other precision-critical fields. Our inclusive submission policy welcomed both archival and non-archival contributions, aimed at fostering interdisciplinary exchange and supporting early-stage and exploratory work.

Prior to the workshop, CHOMPS 2025 hosted a shared task: SHROOM-CAP (Shared-task on Hallucinations and Related Observable Overgeneration Mistakes in Crosslingual Analyses of Publications). SHROOM-CAP invited participants to detect scientific hallucinations in LLM-generated text in a challenging cross-lingual setting spanning high-resource languages (English, Spanish, French, Hindi, Italian) as well as several Indic languages with limited NLP resources (Bengali, Telugu, Malayalam, Gujarati).

In total, we received 24 submissions for the workshop. Of these, three teams that participated in the shared task also submitted system description papers. After review, six submissions were accepted as archival papers and invited four non-archival presentations. This corresponds to a 40% acceptance rate for archival submissions. In addition to these, this volume includes one shared task overview paper and all system description papers.

We are grateful to our invited keynote speakers: Abhilasha Ravichander (University of Washington, USA), Danish Pruthi (IISc Bangalore, India), Khyathi Raghavi Chandu (Mistral AI, USA), and Anna Rogers (IT University of Copenhagen, Denmark). We also extend our thanks to the members of our Panel Discussion. At the time of assembling these proceedings, we were still awaiting final confirmations, and we are grateful to all who agreed to contribute their time and expertise. We are especially grateful to the members of the Program Committee, who served as reviewers and dedicated their time and expertise to ensuring the high quality of the workshop. We hope that this event and the work collected in these

proceedings will spark new collaborations and help pave the way toward more reliable, transparent, and linguistically inclusive language technologies.

The CHOMPS organizers,

Aman Sinha, Raúl Vázquez, Timothee Mickus, Rohit Agarwal, Ioana Buhnila, Patrícia Schmidtová, Federica Gamba, Dilip K. Prasad and Jörg Tiedemann

Program Committee

Program Chairs

Aman Sinha, University of Lorraine, France
Raúl Vázquez, University of Helsinki, Finland
Timothee Mickus, University of Helsinki, Finland
Rohit Agarwal, UiT Tromsø, Norway
Ioana Buhnila, Chosun University, South Korea
Patrícia Schmidtová, Charles University, Prague
Federica Gamba, Charles University, Prague
Dilip K Prasad, UiT Tromsø, Norway
Jörg Tiedemann, University of Helsinki, Finland

Program Committee

Joseph Attieh, University of Helsinki
Loris Bergeron, University of Luxemburg
Patanjali Bhamidipati, IIIT Hyderabad
George Drayson, UCL AI Centre
Fanny Ducel, Université Paris-Saclay
Ondřej Dušek, Charles University
Bryan Eikema, University of Amsterdam
Félix Gaschi, POSOS
Ona de Gibert, University of Helsinki
Jindřich Helcl, Charles University
Aditya Joshi, UNSW
Yash Kankanampati, Université Paris Nord (Paris XIII)
Priyanshu Kumar, Apple
Jindřich Libovický, Charles University
Kristýna Onderková, Charles University
Alessandro Raganato, University Milano-Bicocca
Frederic Sadrieh, Hasso Plattner Institute
Claudio Savelli, Politecnico di Torino
Rohit Saxena, University of Edinburgh
Patricia Schmidtova, Charles University
Vincent Segonne, Université Bretagne Sud
Ondřej Sotolář, Masaryk University
Tarun Tater, University of Stuttgart
Teemu Vahtola, University of Helsinki
Amelie Wuhrl, University of Copenhagen
Zhuohan Xie, MBZUAI
Laura Zanella, POSOS
Xinyu Crystina Zhang, University of Waterloo

Publication Chair

Raúl Vázquez, University of Helsinki, Finland

Invited Speakers

Abhilasha Ravichander, University of Washington, USA
Danish Pruthi, IISc Bangalore, India
Khyathi Raghavi Chandu, Mistral AI, USA
Anna Rogers, IT University of Copenhagen, Denmark

Keynote Talk

Illuminating Generative AI: Mapping Knowledge in Large Language Models

Abhilasha Ravichander
University of Washington, USA
2025-23-12 09:10 –

Abstract: Millions of everyday users are interacting with technologies built with generative AI, such as voice assistants, search engines, and chatbots. While AI-based systems are being increasingly integrated into modern life, they can also magnify risks, inequities, and dissatisfaction when providers deploy unreliable systems. A primary obstacle to having more reliable systems is the opacity of the underlying large language models— we lack a systematic understanding of how models work, where critical vulnerabilities may arise, why they are happening, and how models must be redesigned to address them. In this talk, I will first describe my work in investigating large language models to illuminate when models acquire knowledge and capabilities. Then, I will describe my work on building methods to enable data transparency for large language models, that allows practitioners to make sense of the information available to models. Finally, I will describe work on understanding why large language models produce incorrect knowledge, and implications for building the next generation of responsible AI systems.

Bio: Abhilasha Ravichander is a postdoctoral scholar at the Paul G. Allen Center for Computer Science and Engineering at the University of Washington. Her work focuses on building trustworthy language models by developing rigorous diagnostic techniques for models and datasets, and by creating methods to understand large language models and the principles that govern them.

Keynote Talk

Cultural Misrepresentations in AI-generated Stories

Danish Pruthi
IISc Bangalore, India
2025-23-12 11:30 –

Abstract: TBA

Bio: Danish Pruthi is an Assistant Professor at the Indian Institute of Science (IISc), Bangalore. He received his Ph.D. from the School of Computer Science at Carnegie Mellon University. His research focuses on addressing issues concerning the interpretability of deep learning models, and more recently, in geo-cultural representation in AI and understanding the behavior of Large Language Models.

Keynote Talk

Decoding Multimodal Uncertainty and Reliability in knowledge quadrant

Khyathi Raghavi Chandu
Mistral AI
2025-23-12 13:30 –

Abstract: Ensuring the reliability of vision-language models (VLMs) is crucial for their application in real-world AI contexts, particularly in critical domains where tracing and recovering from errors is challenging. While existing methodologies like selective prediction and image generation have made strides, challenges persist in enabling robust reasoning and accurate predictions under uncertainty. I postulate that the fundamental reason for this gap is not extensively exploring the knowledge quadrant. This talk addresses two key questions: (1) How can we train models to abstain answering unknown-unknowns when uncertain and defer to human judgements? (2) How can we mitigate over-refusals and hallucinations from unknown-knowns without compromising performance? Can we effectively use VLMs and LLMs to serve as agents to enhance performance and certainty in unknown-known conditions? First, I will introduce CertainlyUncertain, a benchmark dataset designed to challenge VLMs with uncertain scenarios. I will demonstrate the empirical improvement of our models in accurate refusals (UNK-VQA, TDIUC) and reducing hallucinations (MM-Hal, POPE) while maintaining general capabilities (VQAv2, VizWiz). Second, I will present our ReCoVERR algorithm, which utilizes vision and language tools as agents to accumulate confidence information during inference, improving coverage by 20% and recall by 25-30%. I will very briefly touch upon our demo on using generator-critic paired agent to construct and critique unseen objects in 3D simulations.
I will conclude by emphasizing that systematically exploring the knowledge quadrant not only enhances the reliability of LLMs and VLMs but also fosters robustness in real-world interactions with error recovery, ensuring that these models can navigate uncertainty with greater confidence and accuracy.

Bio: Khyathi Raghavi Chandu is a AI Research Scientist at Mistral AI. She received her Ph.D. from Carnegie Mellon University. Her research centers on developing and training large-scale models with an expertise on grounded multimodal long-form generation, more recently, practical pathways for building more reliable LLMs, focusing on multimodality.

Keynote Talk

Factuality and Attribution for Large Language Models

Anna Rogers
IT University of Copenhagen, Denmark
2025-23-12 14:30 –

Abstract: This talk addresses the factuality status of generative language model output, and the ongoing impact of language models on the information ecosphere and content economy. I will also discuss the technical and social challenges of providing source attribution via the current LLM interfaces.

Bio: Anna Rogers is a tenured Associate Professor in the Data Science Section at the IT University of Copenhagen, affiliated with the NLPNorth research group. Her research focuses on model analysis and evaluation of natural language understanding systems, with a keen interest in interpretability and robustness of NLP systems based on Large Language Models.

Table of Contents

Program

Sunday, November 9, 2025

09:00 - 09:10 *Opening Remarks*

09:10 - 10:10 *Keynote Talk 1:* ***Abhilasha Ravichander***

10:10 - 10:30 *Lightning poster session*

10:30 - 11:00 *Coffee Break*

11:00 - 11:30 *Poster session*

11:30 - 12:30 *Keynote Talk 2:* ***Danish Pruthi***

12:30 - 13:30 *Lunch Break*

13:30 - 14:30 *Keynote Talk 3:* ***Khyathi Raghavi Chandu***

14:30 - 15:30 *Keynote Talk 4:* ***Anna Rogers***

15:30 - 16:00 *Coffee Break*

16:00 - 16:15 *Shared Task Overview and Lightning Round for Shared Task Papers*

16:15 - 16:55 *Panel Discussion: Trustworthy and Accurate Multilingual Models in Mission-Critical Contexts*

16:55 - 17:00 *Closing Remarks*