

INLG 2025

**Proceedings
of the
18th International Natural Language Generation Conference**

October 29 – November 2, 2025

©2025 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-321-0

Preface

We are excited to present the Proceedings of the 18th International Natural Language Generation Conference (INLG 2025). This year's INLG takes place from October 29 to November 2, 2025 in Hanoi, Vietnam and is organized by the Vietnam Institute for Advanced Study in Mathematics. We would like to thank the local organizing team led by Nguyễn Lê Minh and Nguyễn Thị Minh Huyền; the conference would not be possible without their dedication and hard work.

The INLG conference is the main international forum for the presentation and discussion of research on Natural Language Generation (NLG). This year, we received 127 conference submissions (including 5 from ARR) and 8 demo paper submissions. After a peer review process, 38 long papers, 12 short papers, and 4 demo papers were accepted to the conference and are included in these proceedings. The accepted papers showcase the breadth of NLG research, including work on applications, such as data-to-text tasks, machine translation, and summarization; language model evaluation; and many other topics of interest to the NLG community. We thank Ondřej Dušek for serving as Publication Chair and preparing these proceedings.

We are also excited to present five keynotes, which will discuss the next frontier of AI alignment, social intelligence with LLMs, LLM evaluation in high-stakes, evaluation of safety in LLM outputs, and LLM cooperators for research.

The keynote speakers are:

- Minlie Huang, Tsinghua University, China
- Iryna Gurevych, TU Darmstadt, Germany
- Verena Rieser, Google DeepMind, UK
- Hadas Kotek, Apple, USA
- Michael White, Ohio State University, USA

In continuation of a long tradition that goes back to 2007, INLG is hosting a Generation Challenge, a track of the main conference focused on developing shared tasks for NLG. The track is chaired by Simon Mille and is described in its own proceedings volume.

Four workshops are co-located with the main conference: LLM Reasoning on Medicine: Challenges, Opportunities, and Future, The 1st Workshop for Young Researchers in Natural Language Generation, The 3rd International Workshop of AI Werewolf and Dialog System, and The 11th International Workshop on Vietnamese Language and Speech Processing. INLG is also hosting two tutorials: “Visual Context” and “Large Language Models in Social Science: Methods, Applications, and Ethics.” We also thank Xiaoyong Wei for serving as the Workshop and Tutorial Chair for the conference.

This event is sponsored by Vingroup Innovation Foundation (VINIF – VinBigData). We would like to thank our generous sponsor and we thank Shreyas Sharma for serving as the sponsor Chair for the conference.

We would also like to express our gratitude to the Area Chairs and Program Committee members for their review contributions, and to the SIGGEN representatives – Chenghua Lin, David M. Howcroft, Saad Mahamood, Simon Mille and Patrícia Schmidtová – for sharing their expertise.

Your INLG 2025 program chairs,
Lucie Flek, Shashi Narayan, Lê Hồng Phương, and Jiahuan Pei

Organizing Committee

Program Chairs

Lucie Flek, University of Bonn / Lamarr Institute of Machine Learning and Artificial Intelligence
Shashi Narayan, Google DeepMind
Lê Hồng Phương, VNU Hanoi University of Science
Jiahuan Pei, Vrije Universiteit Amsterdam

Generation Challenge Chair

Simon Mille, ADAPT Research Centre / Dublin City University

Workshop & Tutorial Chair

Xiaoyong Wei, The Hong Kong Polytechnic University

Local Organizer Chairs

Nguyễn Lê Minh, Japan Advanced Institute of Science and Technology
Nguyễn Thị Minh Huyền, VNU Hanoi University of Science

Local Organizers

Nguyễn Việt Cường, Intelligent Integration Co., Ltd. (INT²)
Đỗ Văn Hải, Thuy Loi University
Phan Xuân Hiếu, VNU University of Engineering and Technology
Lê Hoàng Quỳnh, VNU University of Engineering and Technology
Nguyễn Phương Thái, VNU University of Engineering and Technology
Nguyễn Minh Tiến, Hung Yen University of Technology and Education
Nguyễn Thị Thu Trang, Hanoi University of Science and Technology
Trần Đức Vũ, Japan Advanced Institute of Science and Technology

Publication Chair

Ondřej Dušek, Charles University

Sponsor Chair

Shreyas Sharma, aiXplain

SIGGEN Executives

Chenghua Lin, University of Manchester
David M. Howcroft, University of Aberdeen
Saad Mahamood, Shopware
Simon Mille, ADAPT Research Centre / Dublin City University
Patrícia Schmidtová, Charles University

Area Chairs

Reinald Kim Amplayo, Google
Guanyi Chen, Central China Normal University
Ondřej Dušek, Charles University
Sebastian Gehrmann, Bloomberg LP
Kelvin Han, Independent
Rudali Huidrom, ADAPT Research Centre / Dublin City University
Johannes Kiesel, Bauhaus-Universität Weimar
Lara J. Martin, University of Maryland, Baltimore County
Laura Perez-Beltrachini, University of Edinburgh
Ehud Reiter, University of Aberdeen
Fahime Same, trivago N.V.
João Sedoc, New York University
Sina Zarrieß, University of Bielefeld

Program Committee

Rim Abrougui, Aday
Alyssa Allen, The Ohio State University
Miriam Anschütz, Technical University of Munich
Mary-Jane Antia, University of Cape Town
Xin Bai, Central China Normal University
Anya Belz, ADAPT Research Centre / Dublin City University
Nadjet Bouayad-Agha, NLP Scientist
Daniel Braun, University of Marburg
Gordon Briggs, U.S. Naval Research Laboratory
Alberto Bugarín-Diz, University of Santiago de Compostela
Eduardo Calò, Utrecht University
Thiago Castro Ferreira, Federal University of Minas Gerais
Guanyi Chen, Central China Normal University
Chung-Chi Chen, National Institute of Advanced Industrial Science and Technology
Wei-Fan Chen, University of Bonn
Watson Wei Khong Chua, Government Technology Agency of Singapore
Miruna Adriana Clinciu, Edinburgh Centre for Robotics
Hugo Contant, Carnegie Mellon University
Brian Davis, Dublin City University
Simon Dobnik, University of Gothenburg
Shen Dong, University of Glasgow
Pablo Duboue, Textualization Software
Ondřej Dušek, Charles University
Lucie Flek, University of Bonn / Lamarr Institute of Machine Learning and Artificial Intelligence
Albert Gatt, Utrecht University
Ray Groshan, University of Maryland, Baltimore County
Kelvin Han, Independent
Kei Harada, The University of Electro-Communications
Hiroaki Hayashi, Salesforce Research
Arya Honraopatil, University Of Maryland Baltimore County
David M. Howcroft, University of Aberdeen
Nikolai Illykh, University of Gothenburg
Simeon Junker, Bielefeld University
Mihir Kale, Meta

Srinivas Ramesh Kamath, trivago N.V.
 Yoshinobu Kano, Shizuoka University
 Debanjana Kar, IIT Kharagpur
 Zdeněk Kasner, Charles University
 Gary Kazantsev, Bloomberg
 Emiel Krahmer, Tilburg University
 Lea Krause, Vrije Universiteit Amsterdam
 Mateusz Lango, Poznan University of Technology / Charles University
 Guy Lapalme, RALI-DIRO Université de Montréal
 Ashley Lewis, The Ohio State University
 Jin Li, Shenzhen Institute of Advanced Technology / Chinese Academy of Sciences
 Stephan Linzbach, GESIS - Leibniz Institute for the Social Sciences
 Wei-Yun Ma, Academia Sinica
 Saad Mahamood, Shopware
 Zola Mahlaza, University of Cape Town
 Aleksandre Maskharashvili, University of Illinois Urbana-Champaign
 Kathleen F. McCoy, University of Delaware
 David D. McDonald, Smart Information Flow Technologies
 Gonzalo Méndez, Universidad Complutense de Madrid
 Qingyu Meng, Vrije University Amsterdam
 Antonio Valerio Miceli Barone, The University of Edinburgh
 Simon Mille, ADAPT Research Centre / Dublin City University
 Yifan Mo, Vrije Universiteit Amsterdam
 Anna Nikiforovskaya, CNRS/LORIA / Université de Lorraine
 Kristýna Onderková, Charles University
 Suraj Pandey, The Open University
 Siyana Pavlova, Université de Lorraine
 Jiahuan Pei, Vrije Universiteit Amsterdam
 Pablo N Perez De Angelis, Gezie.io
 Minh Vu Pham, IT:U Austria
 Toky Hajatiana Raboanary, University of Cape Town
 Vandana Sreenivasa Rao, Microsoft
 Ehud Reiter, University of Aberdeen
 Philipp Sadler, University of Potsdam
 Fahime Same, trivago N.V.
 Daniel Sanchez, University of Granada
 Sashank Santhanam, University of North Carolina at Charlotte / Apple
 Patrícia Schmidtová, Charles University
 Anastasia Shimorina, Orange
 Judith Sieker, Bielefeld University
 Adarsa Sivaprasad, University of Aberdeen
 Yifei Song, CNRS-LORIA
 Yingjin Song, Utrecht University
 William Eduardo Soto Martinez, LORIA
 Somayajulu Sripada, Arria NLG / University of Aberdeen
 Prerak Srivastava, SAP Labs
 Symon Stevens-Guille, The Ohio State University
 Kristina Striegnitz, Union College
 Yue Su, Vrije Universiteit Amsterdam
 Barkavi Sundararajan, University of Aberdeen
 Jan Svec, Brno University of Technology

Hiroya Takamura, The National Institute of Advanced Industrial Science and Technology
Ekaterina Taktasheva, University of Edinburgh
Marc Tanti, University of Malta
Mariet Theune, University of Twente
Ilias Triantafyllopoulos, New York University
Qingyun Wang, William & Mary
Robert Weißgraeber, AX Semantics
Hugh Mee Wong, Utrecht University
Siwei Wu, Nanjing University of Science & Technology
Xinnuo Xu, Microsoft Research
Bohao Yang, University of Manchester
Kun Zhang, INRIA Saclay / École Polytechnique
Huajian Zhang, Westlake University
Tianyi Zhang, University of Pennsylvania
Ingrid Zukerman, Monash University
Rodrigo de Oliveira, IQVIA
Chris van der Lee, Tilburg University

Best Area Chairs

Ondřej Dušek
Sebastian Gehrmann
Kelvin Han
Laura Perez-Beltrachini
Ehud Reiter
Fahime Same

Best Reviewers

Miriam Anshütz
Nadjet Bouayad-Agha
Chung-Chi Chen
Kathleen F. McCoy
Kelvin Han
David M. Howcroft
Lea Krause
Ashley Lewis
Wei-Yun Ma
Toky Hajatiana Raboanary
Fahime Same
Qingyun Wang

Table of Contents

<i>Enhancing Coherence and Interestingness in Knowledge-Grounded Dialogue Generation</i> Hiroki Onozeki and Michimasa Inaba	1
<i>Evaluating LLM-Generated Versus Human-Authored Responses in Role-Play Dialogues</i> Dongxu Lu, Johan Jeuring and Albert Gatt	20
<i>Human ratings of LLM response generation in pair-programming dialogue</i> Cecilia Domingo, Paul Piwek, Svetlana Stoyanchev, Michel Wermelinger, Kaustubh Adhikari and Rama Sanand Doddipatla	41
<i>Do My Eyes Deceive Me? A Survey of Human Evaluations of Hallucinations in NLG</i> Patricia Schmidtova, Eduardo Calò, Simone Balloccu, Dimitra Gkatzia, Rudali Huidrom, Mateusz Lango, Fahime Same, Vilém Zouhar, Saad Mahamood and Ondrej Dusek	60
<i>Truth or Twist? Optimal Model Selection for Reliable Label Flipping Evaluation in LLM-based Counterfactuals</i> Qianli Wang, Van Bach Nguyen, Nils Feldhus, Luis Felipe Villa-Arenas, Christin Seifert, Sebastian Möller and Vera Schmitt	80
<i>Assessing Semantic Consistency in Data-to-Text Generation: A Meta-Evaluation of Textual, Semantic and Model-Based Metrics</i> Rudali Huidrom, Michela Lorandi, Simon Mille, Craig Thomson and Anya Belz	98
<i>FreshTab: Sourcing Fresh Data for Table-to-Text Generation Evaluation</i> Kristýna Onderková, Ondrej Platek, Zdeněk Kasner and Ondrej Dusek	108
<i>KDA: Knowledge Distillation Adapter for Cross-Lingual Transfer</i> Ta-Bao Nguyen, Nguyen-Phuong Phan, Tung Le and Huy Tien Nguyen	122
<i>ViNumFCR: A Novel Vietnamese Benchmark for Numerical Reasoning Fact Checking on Social Media News</i> Nhi Ngoc Phuong Luong, Anh Thi Lan Le, Tin Van Huynh, Kiet Van Nguyen and Ngan Nguyen	134
<i>Dual Debiasing: Remove Stereotypes and Keep Factual Gender for Fair Language Modeling and Translation</i> Tomasz Limisiewicz, David Mareček and Tomáš Musil	148
<i>Mining Contextualized Visual Associations from Images for Creativity Understanding</i> Ananya Sahu, Amith Ananthram and Kathleen McKeown	165
<i>Analysing Reference Production of Large Language Models</i> Chengzhao Wu, Guanyi Chen, Fahime Same and Tingting HE	182
<i>Live Football Commentary (LFC): A Large-Scale Dataset for Building Football Commentary Generation Models</i> Taiga Someya, Tatsuya Ishigaki and Hiroya Takamura	195
<i>Exploring the Power of Large Language Models for Vietnamese Implicit Sentiment Analysis</i> Huy Gia Luu and Dang Van Thin	202

<i>Towards Trustworthy Lexical Simplification: Exploring Safety and Efficiency with Small LLMs</i> Akio Hayakawa, Stefan Bott and Horacio Saggion	215
<i>Evaluating LLMs' Ability to Understand Numerical Time Series for Text Generation</i> Mizuki Arai, Tatsuya Ishigaki, Masayuki Kawarada, Yusuke Miyao, Hiroya Takamura and Ichiro Kobayashi	232
<i>Can GPT models Follow Human Summarization Guidelines? A Study for Targeted Communication Goals</i> Yongxin Zhou, Fabien Ringeval and François Portet	249
<i>References Matter: Investigating the Impact of Reference Set Variation on Summarization Evaluation</i> Silvia Casola, Yang Janet Liu, Siyao Peng, Oliver Kraus, Albert Gatt and Barbara Plank	274
<i>OpenNLGauge: An Explainable Metric for NLG Evaluation with Open-Weights LLMs</i> Ivan Kartac, Mateusz Lango and Ondrej Dusek	292
<i>Statistical Multicriteria Evaluation of LLM-Generated Text</i> Esteban Garces Arias, Hannah Blocher, Julian Rodemann, Matthias Assenmacher and Christoph Jansen	338
<i>Incorporating Formulaicness in the Automatic Evaluation of Naturalness: A Case Study in Logic-to-Text Generation</i> Eduardo Calò, Guanyi Chen, Elias Stengel-Eskin, Albert Gatt and Kees van Deemter	352
<i>Surprisal reveals diversity gaps in image captioning and different scorers change the story</i> Nikolai Ilinykh and Simon Dobnik	366
<i>Natural Language Translation of Formal Proofs through Informalization of Proof Steps and Recursive Summarization along Proof Structure</i> Seiji Hattori, Takuya Matsuzaki and Makoto Fujiwara	376
<i>German4All – A Dataset and Model for Readability-Controlled Paraphrasing in German</i> Miriam Anshütz, Thanh Mai Pham, Eslam Nasrallah, Maximilian Müller, Cristian-George Craciun and Georg Groh	390
<i>Enhancing Named Entity Translation from Classical Chinese to Vietnamese in Traditional Vietnamese Medicine Domain: A Hybrid Masking and Dictionary-Augmented Approach</i> Nhu Vo Quynh Pham, Uyen Bao Nguyen Phuc, Long Hong Buu Nguyen and Dien Dinh	408
<i>Fine-Tuning, Prompting and RAG for Knowledge Graph-to-Russian Text Generation. How do these Methods generalise to Out-of-Distribution Data?</i> Anna Nikiforovskaya, William Eduardo Soto Martinez, Evan Parker Kelly Chapple and Claire Gardent	419
<i>FinStat2SQL: A Text2SQL Pipeline for Financial Statement Analysis</i> Hung Quang Nguyen, Anh Phuong Trinh, Hung Phan Quoc Mai and Phong Tuan Trinh	449
<i>From Prototypical to Relational: How LLMs Navigate Complex Analogies</i> Mayukh Das and Wolf-Tilo Balke	465
<i>Automated and Context-Aware Code Documentation Leveraging Advanced LLMs</i> Swapnil Sharma Sarker and Tanzina Taher Ifty	486

<i>LogitRouter: a novel Attention variant for reducing Myopic Routing in Mixture of Experts</i>	
Felipe Rodriguez and Marcelo Mendoza	499
<i>Restaurant Menu Categorization at Scale: LLM-Guided Hybrid Clustering</i>	
Seemab Latif, Ashar Mehmood, Selim Turki, Huma Ameer, Ivan Gorban and Faysal Fateh....	511
<i>Taming the Titans: A Survey of Efficient LLM Inference Serving</i>	
Ranran Zhen, Juntao Li, Yixin Ji, Zhenlin Yang, tong liu, Qingrong Xia, Xinyu Duan, Zhefeng Wang, baoxing Huai and Min Zhang	522
<i>Are Multi-Agents the new Pipeline Architecture for Data-to-Text Systems?</i>	
Chinonso Cynthia Osuji, Brian Timoney, Mark Andrade, Thiago Castro Ferreira and Brian Davis	542
<i>Generating Impact and Critique Explanations of Predictions made by a Goal Recognizer</i>	
Jair da Silva Ferreira Junior, Ingrid Zukerman, Enes Makalic, Cecile L. Paris and Mor Vered ..	554
<i>PRICoT: Principle Retrieval and Injection from Inference Successes and Failures for CoT Improvement</i>	
Yudai Yamazaki, Naoto Takeda, Yasutaka Nishimura and Kazushi Ikeda	576
<i>Cognitive Flow: An LLM-Automated Framework for Quantifying Reasoning Distillation</i>	
José Matos, Catarina Silva and Hugo Goncalo Oliveira	596
<i>How (un)faithful are explainable LLM-based NLG metrics?</i>	
Alex Terentowicz, Mateusz Lango and Ondrej Dusek	617
<i>Counterfactual Simulatability of LLM Explanations for Generation Tasks</i>	
Marvin Limpijankit, Yanda Chen, Melanie Subbiah, Nicholas Deas and Kathleen McKeown ..	659
<i>SWI: Speaking with Intent in Large Language Models</i>	
Yuwei Yin, Eunjeong Hwang and Giuseppe Carenini	684
<i>Forecasting Conversation Derailments Through Generation</i>	
Yunfan Zhang, Kathleen McKeown and Smaranda Muresan	699
<i>Can LLMs Help Encoder Models Maintain Both High Accuracy and Consistency in Temporal Relation Classification?</i>	
Adiel Meir and Kfir Bar	716
<i>Benchmarking and Improving LVLMs on Event Extraction from Multimedia Documents</i>	
Fuyu Xing, Zimu Wang, Wei Wang and Haiyang Zhang	734
<i>QCoder Benchmark: Bridging Language Generation and Quantum Hardware through Simulator-Based Feedback</i>	
Taku Mikuriya, Tatsuya Ishigaki, Masayuki Kawarada, Shunya Minami, Tadashi Kadowaki, Yohichi Suzuki, Soshun Naito, Shunya Takada, Takumi Kato, Tamotsu Basseda, Reo Yamada and Hiroya Takamura	743
<i>When LLMs Can't Help: Real-World Evaluation of LLMs in Nutrition</i>	
Karen Jia-Hui Li, Simone Balloccu, Ondrej Dusek and Ehud Reiter	753
<i>Who's Laughing Now? An Overview of Computational Humour Generation and Explanation</i>	
Tyler Loakman, William Thorne and Chenghua Lin	780
<i>Input Matters: Evaluating Input Structure's Impact on LLM Summaries of Sports Play-by-Play</i>	
Barkavi Sundararajan, Somayajulu Sripada and Ehud Reiter	795

<i>Scaling Up Data-to-Text Generation to Longer Sequences: A New Dataset and Benchmark Results for Generation from Large Triple Sets</i>	
Chinonso Cynthia Osuji, Simon Mille, Ornait O’Connell, Thiago Castro Ferreira, Anya Belz and Brian Davis	810
<i>Annotating Hallucinations in Question-Answering using Rewriting</i>	
Xu Liu, Guanyi Chen, Kees van Deemter and Tingting HE	823
<i>Effectiveness of Chain-of-Thought in Distilling Reasoning Capability from Large Language Models</i>	
Cong Thanh Do, Rama Sanand Doddipatla and Kate Knill	833
<i>Face the Facts! Evaluating RAG-based Pipelines for Professional Fact-Checking</i>	
Daniel Russo, Stefano Menini, Jacopo Staiano and Marco Guerini	846