

DSTC 2025

The Twelfth Dialog System Technology Challenge (DSTC12)

Proceedings of the Workshop

August 28, 2025

©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

ISBN 979-8-89176-330-2

Introduction

We are excited to welcome you to DSTC-12, the Twelfth Dialog System Technology Challenge. This year the workshop is being held on August 28th, 2025 at SIGDial 2025.

The DSTC shared tasks have provided common testbeds for the dialog research community since 2013. This year, the program includes two invited talks from Verena Rieser and Milica Gašić, two track presentations, four paper presentations and a panel with Ryuichiro Higashinaka, Laurent Prévot, Tetsuro Takahashi, Milica Gašić.

We had two tracks this year: Track 1: Dialog System Evaluation: Dimensionality, Language, Culture and Safety, organized by John Mendonca, Lining Zhang, Rahul Mallidi, Alon Lavie, Isabel Trancoso, Luis Fernando D’Haro, João Sedoc Track 2: Controllable Conversational Theme Detection Track, organized by Igor Shalyminov, Hang Su, Jake Vincent, Siffi Singh, Jason Cai, James Gung, Raphael Shu, Saab Mansour

The track timelines are: Jan – April 2025 training phase April 2025 testing phase May 2025 results announcement

We are very proud of this years DSTC and would like to thank some key players. We would like to thank our session chairs (Emre Can, Alexandru Coca), our panelist moderator Luis Fernando D’Haro, all our track organizers and a big thank you to all the track participants for their fine work. Finally we would like to thank the wider DSTC community for helping making DSTC a success and helping it to grow for over 10 years.

Behnam Hedayatnia, General Chair

Michel Galley, Publicity Chair

Raghav Gupta, Publication Chair

Zhang Chen and Yun-Nung (Vivian) Chen, Workshop Co-Chairs

Organizing Committee

General Chair

Behnam Hedayatnia, Apple

Workshop Chairs

Vivian Chen, National Taiwan University

Zhang Chen, National Taiwan University, Taiwan

Publication Chair

Raghav Gupta, Google

Publicity Chair

Michel Galley, Microsoft

Program Committee

Program Chairs

Yun-Nung Chen

Michel Galley, Raghav Gupta

Behnam Hedayatnia

Chen Zhang

Reviewers

Namo Bang

Qian Chen, Anoop Cherian, Paul A. Crook

Suvodip Dey

Michel Galley, Kallirroi Georgila

Behnam Hedayatnia, Vojtech Hudecek, Vojtech Hudecek

Seongho Joo

Seokhwan Kim, Sarvesh Kirthivasan

Aditya Nair

Alexandros Papangelis, Baolin Peng

Saurav Sahay, Sashank Santhanam, Harsh Sharma, Prachee Sharma

David Thulke

Bin Wang, George Z Wei

Qi Zhu

Invited Talk
Intentional, Plural, Deep: The Foundations of Beneficial AI Alignment

Verena Rieser
Google Deepmind



August 28, 2025 – Time: 09:45 – 10:25 –

Abstract: We talk constantly about making AI aligned, but we rarely ask the most important questions: aligned to what, and to whom? This keynote deconstructs our current assumptions and proposes that truly beneficial AI requires moving through three distinct depths of alignment. The first is Intentionality: we must stop assuming alignment will emerge on its own and start making deliberate choices about the goals we set. The second is Plurality: we must abandon the search for a single gold standard of human values and engineer systems that thrive on diversity. The third and most profound is Depth: we must push beyond optimizing for clicks and convenience and instead align AI with the complex, often contradictory, nature of long-term human well-being. This is a framework for the next frontier of AI — one that builds technology to support our human nature, not exploit it.

Bio: Verena Rieser is a Senior Staff Research Scientist at Google DeepMind, where she founded and lead the VOICES (Voices-of-all in alignment) team. Their mission is to ensure that powerful AI models like Gemini are developed responsibly, making them safe and genuinely useful for diverse communities worldwide. Verena’s career has been driven by a fascination with conversational AI. Verena has pioneered research in dialogue systems and natural language generation, always with a focus on applications that create societal benefit. Before joining Google, Verena was a full professor directing the NLP lab at Heriot-Watt University and held a Royal Society Leverhulme Senior Research Fellowship

Invited Talk

Dimensions of intelligence

Milica Gašić

Heinrich-Heine-University Düsseldorf



August 28, 2025 – Time: 15:00 – 15:40 –

Abstract: Large language models (LLMs) have transformed the area of artificial intelligence, achieving or surpassing human performance in a number of natural language processing tasks. Despite this tremendous success, they lack the ability to model multi-turn goal-directed conversation, they are largely uncalibrated and there is little insight into the way they operate. In this talk, I will present (1) multi-turn optimisation which integrates emotion, (2) confidence-based learning based on insights from human psychology and (3), in the direction of explainability, observations from topological data analysis applied to LLMs. To conclude, I will hypothesise how combining ideas from LLMs and task-oriented systems can lead to conversational agents encompassing a large spectrum of desired properties.

Bio: Milica Gašić is a Professor in Dialog Systems and Machine Learning at Heinrich Heine University. Her research focuses on fundamental questions of human-computer dialogue modelling and lie in the intersection of Natural Language Processing and Machine Learning. She is a recipient of the European Research Council Starting Grant and the Alexander von Humboldt Sofja Kovalevskaja Award. Prof. Gašić is a member of the International Scientific Advisory Board of DFKI, a member of ACL, a member of ELLIS and a senior member of IEEE. She served as the vice-president of SIGDIAL 2021-2025.

Table of Contents

<i>Neural Models and Language Model Prompting for the Multidimensional Evaluation of Open-Ended Conversations</i>	
Michelle Elizabeth, Alicja Kasicka, Natalia Krawczyk, Magalie Ochs, Gwéno�� Lecorv��, Justyna Gromada and Lina M. Rojas-Barahona	1
<i>CATCH: A Controllable Theme Detection Framework with Contextualized Clustering and Hierarchical Generation</i>	
Rui Ke, Jiahui Xu, Kuang Wang, Shenghao Yang, Feng Jiang and Haizhou Li	17
<i>Overview of Dialog System Evaluation Track: Dimensionality, Language, Culture and Safety at DSTC 12</i>	
John Mendon��a, Lining Zhang, Rahul Mallidi, Alon Lavie, Isabel Trancoso, Luis Fernando D’Haro and Jo��o Sedoc	27
<i>The Limits of Post-hoc Preference Adaptation: A Case Study on DSTC12 Clustering</i>	
Jihyun Lee and Gary Lee	36
<i>KSTC: Keyphrase-driven Sentence embedding and Task independent prompting for filling slot in the Generation of theme label</i>	
Sua Kim, Taeyoung Jeong, Seokyoung Hong, Seongjun Kim, Jeongpil Lee, Du-Seong Chang and Myoung-Wan Koo	44
<i>Controllable Conversational Theme Detection Track at DSTC 12</i>	
Igor Shalymov, Hang Su, Jake W. Vincent, Siffi Singh, Jason Cai, James Gung, Raphael Shu and Saab Mansour	74