

SIGDIAL 2023



**The 24th Meeting of the  
Special Interest Group on Discourse and  
Dialogue**



**Proceedings of the Conference**

September 11 - 15, 2023  
Prague, Czechia

©2023 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](mailto:acl@aclweb.org)

ISBN 979-8-89176-028-8

## Preface

We are glad to pen the first few words for the proceedings of SIGDIAL 2023, the 24rd Annual Meeting of the Special Interest Group on Discourse and Dialogue. The SIGDIAL conference is a premier publication venue for research in discourse and dialogue. This year the conference is organized together with the conference on International Natural Language Generation (INLG). The format is hybrid with most participants and presenters in-person. Zoom was used for remote presentations and Discord was used as a communication platform for both remote and local participants.

The joint SIGDIAL-INLG 2023 took place on September 11-15, 2023 in Prague, Czech Republic at OREA Hotel Pyramida. The joint conference was collocated with five full-day and one half-day workshops and one satellite event on September 11-12:

- Taming Large Language Models: Controllability in the era of Interactive Assistants
- Workshop on Multimodal, Multilingual Natural Language Generation and Multilingual WebNLG Challenge
- The 1st Workshop on Counter Speech for Online Abuse
- The Eleventh Dialog System Technology Challenge
- PracticalD2T: 1st Workshop on Practical LLM-assisted Data-to-Text Generation
- Connecting multiple disciplines to AI techniques in interaction-centric autism research and diagnosis (half-day workshop)
- The 2023 Young Researchers' Roundtable on Spoken Dialog Systems (YRRSDS 2023) was held as a satellite event

We thank the organizers of the workshops. We are grateful for their coordination with the main conference.

SIGDIAL received 136 submissions this year, comprising 87 long papers, 42 short papers, and 7 demo descriptions. We have aimed to develop a broad, varied program spanning the many positively-rated papers identified by the review process. We therefore accepted 61 papers in total: 41 long papers (47%), 16 short papers (38%), and 4 demo descriptions, for an overall acceptance rate of 45.5%. In addition, we considered 7 articles from Dialogue and Discourse journal and selected 3 for poster presentations.

SIGDIAL-INLG 2023 jointly featured 4 keynotes (one of them presented remotely), a panel discussion, and a joint virtual poster/demo session. The SIGDIAL 2023 program featured 5 oral sessions with 24 talks. The presented topics included Analysis of discourse and dialogue, LLM for dialogue, Dialogue modeling and evaluation, Language understanding and multimodality, and Topics in open-domain dialogue (arguments, opinions, empathy).

7 of the talks were presented virtually and were evenly distributed among the oral sessions. The format of the talks was a 15 minute presentation with 5 minutes for Q&A. The conference had two in-person poster-demo sessions featuring a total of 27 poster and 2 demo presentations. A virtual joint SIGDIAL-INLG poster/demo session was held on Discord during the conference and featured 8 posters and 2 demos from SIGDIAL submissions.

In organizing this hybrid in-person/ remote conference, we have tried to maintain as much of the spirit of a fully in-person conference as possible, allowing opportunities for questions and discussion both from in-person and remote audiences. Online participants were able to ask questions using the Discord platform which also featured a channel for online discussions.

We had 131 reviewers and 13 Senior Program Committee (SPC) members, who were each responsible for 9-11 papers, leading the discussion process and also contributing with meta-reviews. Each submission was assigned to an SPC member and received at least three reviews. Decisions carefully considered the original reviews, meta-reviews, and discussions among reviewers facilitated by the SPCs. We are immensely grateful to the members of the Program Committee and Senior Program Committee for their efforts in providing excellent, thoughtful reviews of the large number of submissions. Their contributions have been essential to selecting the accepted papers and providing a high-quality technical program for the conference.

A conference of this scale requires the energy, guidance, and contributions of many parties, and we would like to take this opportunity to thank and acknowledge them all. We thank our four keynote speakers, Emmanuel Dupoux (Ecole des Hautes Etudes en Sciences Sociales), Ryan Lowe (OpenAI), Barbara Di Eugenio (University of Illinois Chicago), and Elena Simperl (King's College London) for their inspiring talks on "Textless NLP: towards language processing from raw audio", "Aligning ChatGPT: past, present, and future", "Engaging the Patient in Healthcare: Summarization and Interaction", and "Knowledge graph use cases in natural language generation", respectively.

Ryan Lowe's talk was followed by a panel discussion on 'Social Impact of LLMs'. We thank the panel chair David Traum and the Panelists: Malihe Alikhani, Maria Keet, Ryan Lowe, and Ehud Reiter for engaging discussion on this important topic.

SIGDIAL 2023 was made possible by the dedication and hard work of our community, and we are indebted to many. The hybrid nature (in-person and remote), the collocation with the INLG and seven workshops put additional burden on the organization process. The conference would not have been possible without the advice and support of the SIGDIAL board, particularly Gabriel Skantze and Milica Gasic as well as Emiel van Miltenburg and Dave Howcroft who helped coordination between the collocated events.

The tireless work by the local organizing team led by Ondřej Dušek who was involved in countless discussions prior and during the conference coordinating SIGDIAL, INLG, and collocated workshops. We thank the local team who ensured that the conference ran very smoothly, and was enjoyed greatly by all participants. Without that team, there would not have been a conference.

Special thanks go to Zdenek Kasner and Ondrej Platek for their tireless efforts in managing the website with timely updates, and to the team handling various online aspects of participation: Ondrej Platek, Patricia Schmidtova, Dave Howcroft. We would like to thank Patricia Schmidtova, Mateusz Lango and Simone Balloccu for further help with conference preparation. We are grateful to Souro Mukherjee, Kirill Semenov, Nalin Kumar, and Peter Polák, as well as Zdenek Kasner, Ondrej Platek, Patricia Schmidtova, Simone Balloccu, and Mateusz Lango again, for support with the registration, A/V and all other local organizing tasks. Many thanks also go to Jan Hajič for his support, and especially to Anna Kotěšovicová for making all the local arrangements possible. We would also like to thank the sponsorship chair Ramesh Manuvinakurike, who brought to the conference an impressive panel of conference sponsors. We gratefully acknowledge the support of our sponsors: LivePerson (Platinum), LuxAI (Platinum), Apple (Gold), Furhat Robotics (Silver), AX Semantics (Bronze), and Bloomberg (Bronze). In addition, we thank Malihe Alikhani, the publication chair, and Casey Kennington, the mentoring chair for their dedicated service.

Finally, it was our great pleasure to welcome you physically and remotely to the conference. We hope that we have provided an enriching and productive experience at the joint SIGDIAL-INLG 2023.

Svetlana Stoyanchev, Shafiq Joty, Program Co-Chairs  
David Schlangen, General Chair

# Organizing Committee

## General chair:

David Schlangen, University of Potsdam

## Program chair 1:

Svetlana Stoyanchev, Toshiba Cambridge Research Laboratory

## Program chair 2:

Shafiq Joty, Nanyang Technological University / Salesforce Research

## Sponsorship chair:

Ramesh Manuvinakurike, Intel Labs

## Local chair:

Ondrej Dusek, Charles University Prague

## Mentoring chair:

Casey Kennington, Boise State University

## Publication chair:

Malihe Alikhani, Northeastern University

## Program Committee:

Gavin Abercrombie, Heriot Watt University  
Tazin Afrin, Educational Testing Service  
Ron Artstein, USC Institute for Creative Technologies  
Katherine Atwell, Northeastern University  
M Saiful Bari, Nanyang Technological University  
Timo Baumann, Ostbayerische Technische Hochschule Regensburg  
Nitu Bharati, TheOpenUniversity  
Nate Blaylock, Canary Speech  
Johan Boye, KTH  
Kristy Boyer, University of Florida  
Chloé Braud, IRIT, CNRS  
Hendrik Buschmeier, Bielefeld University  
Justine Cassell, Carnegie Mellon University and Inria Paris  
Senthil Chandramohan, Staples  
Lin Chen, Engineering Manager, Meta Platform Inc.  
Derek Chen, Columbia University  
Hailin Chen, NTU  
Zhiyu Chen, Meta  
Jinho D. Choi, Emory University  
Paul Crook, Meta  
Heriberto Cuayahuitl, University of Lincoln  
Vera Demberg, Saarland University  
Nina Dethlefs, University of Hull  
David DeVault, Anticipant Speech, Inc.

Barbara Di Eugenio, University of Illinois at Chicago  
Bosheng Ding, Nanyang Technological University  
Rama Sanand Doddipatla, Toshiba Cambridge Research Laboratory  
Cecilia Domingo, TheOpenUniversity  
Ondrej Dusek, Charles University  
Alex Fabbri, Salesforce AI Research  
Younna Farag, University of Cambridge  
Shutong Feng, Heinrich-Heine-Universität Düsseldorf  
Elisa Ferracane, Abridge AI, Inc.  
Milica Gasic, Heinrich Heine University Duesseldorf  
Kallirroi Georgila, University of Southern California Institute for Creative Technologies  
Alborz Geramifard, Facebook AI  
Felix Gervits, US Army Research Laboratory  
Tirthankar Ghosal, Oak Ridge National Laboratory  
Jonathan Ginzburg, Université Paris Cité  
Akhilesh Deepak Gotmare, Salesforce Research  
Venkata Subrahmanyan Govindarajan, University of Texas at Austin  
Yulia Grishina, Amazon  
David Gros, University of California - Davis  
Prakhar Gupta, Carnegie Mellon University  
Joakim Gustafson, KTH  
Dilek Hakkani-Tur, Amazon Alexa AI  
Devamanyu Hazarika, Amazon  
Jie He, University of Edinburgh  
Larry Heck, Georgia Institute of Technology  
Behnam Hedayatnia, Amazon  
Ryuichiro Higashinaka, Nagoya University/NTT  
Yufang Hou, IBM Research  
Julian Hough, Swansea University  
David M. Howcroft, Edinburgh Napier University  
Christine Howes, University of Gothenburg  
Ruihong Huang, Texas A&M University  
Michimasa Inaba, The University of Electro-Communications  
Mert Inan, Northeastern University  
Koji Inoue, Kyoto University  
Di Jin, Amazon  
Prathyusha Jwalapuram, Rakuten  
Tatsuya Kawahara, Kyoto University  
Simon Keizer, Toshiba Europe Ltd  
Seokhwan Kim, Amazon Alexa AI  
Kazunori Komatani, Osaka University  
Philippe Laban, Salesforce Research  
Mateusz Lango, Poznan University of Technology  
Staffan Larsson, University of Gothenburg  
Kornel Laskowski, Carnegie Mellon University  
Fabrice Lefèvre, Avignon Univ.  
Hsien-chin Lin, Heinrich Heine University  
Zhaojiang Lin, Meta  
Pierre Lison, Norwegian Computing Centre  
Yang Janet Liu, Georgetown University  
Linlin Liu, Nanyang Technological University

Eduardo Lleida Solano, University of Zaragoza  
 Nurul Lubis, Heinrich Heine University  
 Ramesh Manuvinakurike, Intel labs  
 Alessandro Mazzei, Università degli Studi di Torino  
 Teruhisa Misu, Honda Research Institute USA  
 Anhad Mohananey, Google  
 Tasnim Mohiuddin, Nanyang Technological University  
 Han Cheol Moon, Nanyang Technological University  
 Roger Moore, University of Sheffield  
 Sourabrata Mukherjee, Charles University  
 Philippe Muller, IRIT, University of Toulouse  
 Satoshi Nakamura, Nara Institute of Science and Technology  
 Mikio Nakano, C4A Research Institute, Inc.  
 Anna Nedoluzhko, Charles University in Prague  
 Douglas O'Shaughnessy, INRS-EMT (Univ. of Quebec)  
 Suraj Pandey, The Open University  
 Rebecca Passonneau, The Pennsylvania State University  
 Paul Piwek, The Open University  
 Ondrej Plátek, Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal  
 and Applied Linguistics  
 Massimo Poesio, Queen Mary University of London  
 Stephen Pulman, Apple Inc.  
 Matthew Purver, Queen Mary University of London  
 Kun Qian, Columbia University  
 Chengwei Qin, Nanyang Technological University  
 Liang Qiu, Amazon Alexa AI  
 Vikram Ramanarayanan, University of California, San Francisco  
 Hannah Rashkin, Google Research  
 Mathieu Ravaut, Nanyang Technological University  
 Ehud Reiter, University of Aberdeen  
 Giuseppe Riccardi, University of Trento  
 Antonio Roque, Soar Technology  
 Saurav Sahay, Intel Labs  
 Sakriani Sakti, Japan Advanced Institute of Science and Technology / Nara Institute of Science and  
 Technology (JAIST/NAIST)  
 Chinnadhurai Sankar, SliceX AI  
 Ruhi Sarikaya, Amazon  
 David Schlangen, University of Potsdam  
 Ethan Selfridge, LivePerson  
 Samira Shaikh, University of North Carolina at Charlotte  
 Anthony Sicilia, Northeastern University  
 Gabriel Skantze, KTH Speech Music and Hearing  
 Georg Stemmer, Intel Corp.  
 Amanda Stent, Colby College  
 Matthew Stone, Rutgers University  
 Carl Strathearn, Edinburgh Napier University  
 Kristina Striegnitz, Union College  
 Hiroaki Sugiyama, NTT Communication Science Labs.  
 Alessandro Suglia, Heriot-Watt University  
 António Teixeira, DETI/IEETA, University of Aveiro  
 Megh Thakkar, BITS Pilani

Takenobu Tokunaga, Tokyo Institute of Technology  
Stefan Ultes, University of Bamberg  
Carel van Niekerk, Heinrich Heine University  
David Vandyke, Apple  
Renato Vukovic, Heinrich Heine University Düsseldorf  
Hsin-Min Wang, Academia Sinica  
Yi-Chia Wang, Facebook AI  
Nigel Ward, University of Texas at El Paso  
Michael White, The Ohio State University  
Qingyang Wu, Columbia University  
Chien-Sheng Wu, Salesforce  
Deyi Xiong, Tianjin University  
Yuxiao Ye, University of Cambridge  
Koichiro Yoshino, Guardian Robot Project (GRP) RIKEN, Nara Institute of Science and Technology  
Yanchao Yu, Edinburgh Napier University  
Hongli Zhan, University of Texas at Austin  
Tiancheng Zhao, Binjiang Institute of Zhejiang University  
Ruochen Zhao, Nanyang Technological University  
Mingyang Zhou, Post Doctoral Research Scientist at Columbia University  
Ingrid Zukerman, Monash University

**Invited Speakers:**

Barbara Di Eugenio, University of Illinois Chicago  
Emmanuel Dupoux, Ecole des Hautes Etudes en Sciences Sociales (EHESS) / Meta AI Labs  
Elena Simperl, King's College London  
Ryan Lowe, OpenAI



## Table of Contents

<i>Sources of Noise in Dialogue and How to Deal with Them</i> Derek Chen and Zhou Yu .....	1
<i>Investigating Explicitation of Discourse Connectives in Translation using Automatic Annotations</i> Frances Yung, Merel Scholman, Ekaterina Lapshinova-Koltunski, Christina Pollkläsener and Vera Demberg .....	21
<i>What's Hard in English RST Parsing? Predictive Models for Error Analysis</i> Yang Janet Liu, Tatsuya Aoyama and Amir Zeldes .....	31
<i>Grounded Complex Task Segmentation for Conversational Assistants</i> Rafael Ferreira, David Semedo and Joao Magalhaes .....	43
<i>A Statistical Approach for Quantifying Group Difference in Topic Distributions Using Clinical Discourse Samples</i> Grace Lawley, Peter A. Heeman, Jill K. Dolata, Eric Fombonne and Steven Bedrick .....	55
<i>OpinionConv: Conversational Product Search with Grounded Opinions</i> Vahid Sadiri Javadi, Martin Potthast and Lucie Flek .....	66
<i>Dial-M: A Masking-based Framework for Dialogue Evaluation</i> Suvodip Dey and Maunendra Sankar Desarkar .....	77
<i>From Chatter to Matter: Addressing Critical Steps of Emotion Recognition Learning in Task-oriented Dialogue</i> Shutong Feng, Nurul Lubis, Benjamin Ruppik, Christian Geishauser, Michael Heck, Hsien-chin Lin, Carel van Niekerk, Renato Vukovic and Milica Gasic .....	85
<i>Analyzing Differences in Subjective Annotations by Participants and Third-party Annotators in Multi-modal Dialogue Corpus</i> Kazunori Komatani, Ryu Takeda and Shogo Okada .....	104
<i>Frame-oriented Summarization of Argumentative Discussions</i> Shahbaz Syed, Timon Ziegenbein, Philipp Heinisch, Henning Wachsmuth and Martin Potthast .....	114
<i>Towards Multilingual Automatic Open-Domain Dialogue Evaluation</i> John Mendonca, Alon Lavie and Isabel Trancoso .....	130
<i>Dialog Action-Aware Transformer for Dialog Policy Learning</i> Huimin Wang, Wai Chung Kwan and Kam-Fai Wong .....	142
<i>The Wizard of Curiosities: Enriching Dialogues with Fun Facts</i> Frederico Vicente, Rafael Ferreira, David Semedo and Joao Magalhaes .....	149
<i>The Road to Quality is Paved with Good Revisions: A Detailed Evaluation Methodology for Revision Policies in Incremental Sequence Labelling</i> Brielen Madureira, Patrick Kahardipraja and David Schlangen .....	156
<i>The effect of conversation type on entrainment: Evidence from laughter</i> Bogdan Ludusan and Petra Wagner .....	168

<i>'What are you referring to?' Evaluating the Ability of Multi-Modal Dialogue Models to Process Clarificational Exchanges</i>	
Javier Chiyah-Garcia, Alessandro Suglia, Arash Eshghi and Helen Hastie . . . . .	175
<i>PGTask: Introducing the Task of Profile Generation from Dialogues</i>	
Rui Ribeiro, Joao Paulo Carvalho and Luisa Coheur . . . . .	183
<i>Question Generation to Elicit Users' Food Preferences Considering the Semantic Content</i>	
Jie Zeng, Yukiko Nakano and Tatsuya Sakato . . . . .	190
<i>Roll Up Your Sleeves: Working with a Collaborative and Engaging Task-Oriented Dialogue System</i>	
Lingbo Mo, Shijie Chen, Ziru Chen, Xiang Deng, Ashley Lewis, Sunit Singh, Samuel Stevens, Chang-You Tai, Zhen Wang, Xiang Yue, Tianshu Zhang, Yu Su and Huan Sun . . . . .	197
<i>Leveraging Large Language Models for Automated Dialogue Analysis</i>	
Sarah E. Finch, Ellie S. Paek and Jinho D. Choi . . . . .	202
<i>Are Large Language Models All You Need for Task-Oriented Dialogue?</i>	
Vojtěch Hudeček and Ondrej Dusek . . . . .	216
<i>Multi-party Goal Tracking with LLMs: Comparing Pre-training, Fine-tuning, and Prompt Engineering</i>	
Angus Addlesee, Weronika Sieińska, Nancie Gunson, Daniel Hernandez Garcia, Christian Dondrup and Oliver Lemon . . . . .	229
<i>ChatGPT vs. Crowdsourcing vs. Experts: Annotating Open-Domain Conversations with Speech Functions</i>	
Lidiia Ostyakova, Veronika Smilga, Kseniia Petukhova, Maria Molchanova and Daniel Kornev . . . . .	242
<i>DiactTOD: Learning Generalizable Latent Dialogue Acts for Controllable Task-Oriented Dialogue Systems</i>	
Qingyang Wu, James Gung, Raphael Shu and Yi Zhang . . . . .	255
<i>Approximating Online Human Evaluation of Social Chatbots with Prompting</i>	
Ekaterina Svikhnushina and Pearl Pu . . . . .	268
<i>Dialogue Response Generation Using Completion of Omitted Predicate Arguments Based on Zero Anaphora Resolution</i>	
Ayaka Ueyama and Yoshinobu Kano . . . . .	282
<i>Syndicom: Improving Conversational Commonsense with Error-Injection and Natural Language Feedback</i>	
Christopher Richardson and Larry Heck . . . . .	297
<i>"What do others think?": Task-Oriented Conversational Modeling with Subjective Knowledge</i>	
Chao Zhao, Spandana Gella, Seokhwan Kim, Di Jin, Devamanyu Hazarika, Alexandros Papangelis, Behnam Hedayatnia, Mahdi Namazifar, Yang Liu and Dilek Hakkani-Tur . . . . .	309
<i>UD_Japanese-CEJC: Dependency Relation Annotation on Corpus of Everyday Japanese Conversation</i>	
Mai Omura, Hiroshi Matsuda, Masayuki Asahara and Aya Wakasa . . . . .	324
<i>Unravelling Indirect Answers to Wh-Questions: Corpus Construction, Analysis, and Generation</i>	
Zulpiye Yusupujang and Jonathan Ginzburg . . . . .	336
<i>A New Dataset for Causality Identification in Argumentative Texts</i>	
Khalid Al Khatib, Michael Voelske, Anh Le, Shahbaz Syed, Martin Potthast and Benno Stein . . . . .	349

<i>Controllable Generation of Dialogue Acts for Dialogue Systems via Few-Shot Response Generation and Ranking</i>	
Angela Ramirez, Kartik Agarwal, Juraj Juraska, Utkarsh Garg and Marilyn Walker . . . . .	355
<i>Reference Resolution and New Entities in Exploratory Data Visualization: From Controlled to Unconstrained Interactions with a Conversational Assistant</i>	
Abari Bhattacharya, Abhinav Kumar, Barbara Di Eugenio, Roderick Tabalba, Jillian Aurisano, Veronica Grosso, Andrew Johnson, Jason Leigh and Moira Zellner . . . . .	370
<i>CONVERSER: Few-shot Conversational Dense Retrieval with Synthetic Data Generation</i>	
Chao-Wei Huang, Chen-Yu Hsu, Tsu-Yuan Hsu, Chen-An Li and Yun-Nung Chen . . . . .	381
<i>Speaker Role Identification in Call Centre Dialogues: Leveraging Opening Sentences and Large Language Models</i>	
Minh-Quoc Nghiem, Nichola Roberts and Dmitry Sityaev . . . . .	388
<i>Synthesising Personality with Neural Speech Synthesis</i>	
Shilin Gao, Matthew P. Aylett, David A. Braude and Catherine Lai . . . . .	393
<i>Prompting, Retrieval, Training: An exploration of different approaches for task-oriented dialogue generation</i>	
Gonçalo Raposo, Luisa Coheur and Bruno Martins . . . . .	400
<i>Bootstrapping a Conversational Guide for Colonoscopy Prep</i>	
Pulkit Arya, Madeleine Bloomquist, SUBHANKAR CHAKRABORTY, Andrew Perrault, William Schuler, Eric Fosler-Lussier and Michael White . . . . .	413
<i>Applying Item Response Theory to Task-oriented Dialogue Systems for Accurately Determining User’s Task Success Ability</i>	
Ryu Hirai, Ao Guo and Ryuichiro Higashinaka . . . . .	421
<i>An Open-Domain Avatar Chatbot by Exploiting a Large Language Model</i>	
Takato Yamazaki, Tomoya Mizumoto, Katsumasa Yoshikawa, Masaya Ohagi, Toshiki Kawamoto and Toshinori Sato . . . . .	428
<i>Learning Multimodal Cues of Children’s Uncertainty</i>	
Qi Cheng, Mert Inan, Rahma Mbarki, Grace Grmek, Theresa Choi, Yiming Sun, Kimele Persaud, Jenny Wang and Malihe Alikhani . . . . .	433
<i>Grounding Description-Driven Dialogue State Trackers with Knowledge-Seeking Turns</i>	
Alexandru Coca, Bo-Hsiang Tseng, Jinghong Chen, Weizhe Lin, Weixuan Zhang, Tisha Anders and Bill Byrne . . . . .	444
<i>Resolving References in Visually-Grounded Dialogue via Text Generation</i>	
Bram Willemsen, Livia Qian and Gabriel Skantze . . . . .	457
<i>Slot Induction via Pre-trained Language Model Probing and Multi-level Contrastive Learning</i>	
Hoang Nguyen, Chenwei Zhang, Ye Liu and Philip Yu . . . . .	470
<i>The timing bottleneck: Why timing and overlap are mission-critical for conversational user interfaces, speech recognition and dialogue systems</i>	
Andreas Liesenfeld, Alianda Lopez and Mark Dingemans . . . . .	482
<i>Enhancing Task Bot Engagement with Synthesized Open-Domain Dialog</i>	
Miaoran Li, Baolin Peng, Michel Galley, Jianfeng Gao and Zhu (Drew) Zhang . . . . .	496

<i>Enhancing Performance on Seen and Unseen Dialogue Scenarios using Retrieval-Augmented End-to-End Task-Oriented System</i>	
Jianguo Zhang, Stephen Roller, Kun Qian, Zhiwei Liu, Rui Meng, Shelby Heinecke, Huan Wang, silvio savarese and Caiming Xiong .....	509
<i>Transformer-based Multi-Party Conversation Generation using Dialogue Discourse Acts Planning</i>	
Alexander Chernyavskiy and Dmitry Ilvovsky .....	519
<i>Incorporating Annotator Uncertainty into Representations of Discourse Relations</i>	
S. Magalí López Cortez and Cassandra L. Jacobs .....	530
<i>Investigating the Representation of Open Domain Dialogue Context for Transformer Models</i>	
Vishakh Padmakumar, Behnam Hedayatnia, Di Jin, Patrick Lange, Seokhwan Kim, Nanyun Peng, Yang Liu and Dilek Hakkani-Tur .....	538
<i>C3: Compositional Counterfactual Contrastive Learning for Video-grounded Dialogues</i>	
Hung Le, Nancy Chen and Steven C.H. Hoi .....	548
<i>No that's not what I meant: Handling Third Position Repair in Conversational Question Answering</i>	
Vevake Balaraman, Arash Eshghi, Ioannis Konstas and Ioannis Papaioannou .....	562
<i>When to generate hedges in peer-tutoring interactions</i>	
Alafate Abulimiti, Chloé Clavel and Justine Cassell .....	572
<i>PaperPersiChat: Scientific Paper Discussion Chatbot using Transformers and Discourse Flow Management</i>	
Alexander Chernyavskiy, Max Bregeda and Maria Nikiforova .....	584
<i>FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions</i>	
Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afil Farhan, Birthe Nettet, Meriam Moujahid, Tanvi Dinkar, Verena Rieser and Oliver Lemon .....	588
<i>Towards Breaking the Self-imposed Filter Bubble in Argumentative Dialogues</i>	
Annalena Aicher, Daniel Kornmueller, Yuki Matsuda, Stefan Ultes, Wolfgang Minker and Keiichi Yasumoto .....	593
<i>The Open-domain Paradox for Chatbots: Common Ground as the Basis for Human-like Dialogue</i>	
Gabriel Skantze and A. Seza Doğruöz .....	605
<i>MERCY: Multiple Response Ranking Concurrently in Realistic Open-Domain Conversational Systems</i>	
Sarik Ghazarian, Behnam Hedayatnia, Di Jin, Sijia Liu, Nanyun Peng, Yang Liu and Dilek Hakkani-Tur .....	615
<i>Empathetic Response Generation for Distress Support</i>	
Anuradha Welivita, Chun-Hung Yeh and Pearl Pu .....	632
<i>Reasoning before Responding: Integrating Commonsense-based Causality Explanation for Empathetic Response Generation</i>	
Yahui Fu, Koji Inoue, Chenhui Chu and Tatsuya Kawahara .....	645

# Conference Program

Wednesday September 13, 2023

09:00–09:15 Opening

09:15–10:15 Keynote: Textless NLP: towards language processing from raw audio

10:15–10:45 Coffee Break

10:45–12:30 Oral Session 1: Analysis of discourse and dialogue

*Sources of Noise in Dialogue and How to Deal with Them*

Derek Chen and Zhou Yu

*Investigating Explicitation of Discourse Connectives in Translation using Automatic Annotations*

Frances Yung, Merel Scholman, Ekaterina Lapshinova-Koltunski, Christina Pollkläsener and Vera Demberg

*What's Hard in English RST Parsing? Predictive Models for Error Analysis*

Yang Janet Liu, Tatsuya Aoyama and Amir Zeldes

*Grounded Complex Task Segmentation for Conversational Assistants*

Rafael Ferreira, David Semedo and Joao Magalhaes

*A Statistical Approach for Quantifying Group Difference in Topic Distributions Using Clinical Discourse Samples*

Grace Lawley, Peter A. Heeman, Jill K. Dolata, Eric Fombonne and Steven Bedrick

Wednesday September 13, 2023 (continued)

12:30–13:30 Lunch

13:30–15:10 Poster Session 1

*OpinionConv: Conversational Product Search with Grounded Opinions*

Vahid Sadiri Javadi, Martin Potthast and Lucie Flek

*Dial-M: A Masking-based Framework for Dialogue Evaluation*

Suvodip Dey and Maunendra Sankar Desarkar

*From Chatter to Matter: Addressing Critical Steps of Emotion Recognition Learning in Task-oriented Dialogue*

Shutong Feng, Nurul Lubis, Benjamin Ruppik, Christian Geishauser, Michael Heck, Hsien-chin Lin, Carel van Niekerk, Renato Vukovic and Milica Gasic

*Analyzing Differences in Subjective Annotations by Participants and Third-party Annotators in Multimodal Dialogue Corpus*

Kazunori Komatani, Ryu Takeda and Shogo Okada

*Frame-oriented Summarization of Argumentative Discussions*

Shahbaz Syed, Timon Ziegenbein, Philipp Heinisch, Henning Wachsmuth and Martin Potthast

*Towards Multilingual Automatic Open-Domain Dialogue Evaluation*

John Mendonca, Alon Lavie and Isabel Trancoso

*Dialog Action-Aware Transformer for Dialog Policy Learning*

Huimin Wang, Wai Chung Kwan and Kam-Fai Wong

*The Wizard of Curiosities: Enriching Dialogues with Fun Facts*

Frederico Vicente, Rafael Ferreira, David Semedo and Joao Magalhaes

*The Road to Quality is Paved with Good Revisions: A Detailed Evaluation Methodology for Revision Policies in Incremental Sequence Labelling*

Brielen Madureira, Patrick Kahardipraja and David Schlangen

*The effect of conversation type on entrainment: Evidence from laughter*

Bogdan Ludusan and Petra Wagner

**Wednesday September 13, 2023 (continued)**

*'What are you referring to?' Evaluating the Ability of Multi-Modal Dialogue Models to Process Clarificational Exchanges*

Javier Chiyah-Garcia, Alessandro Suglia, Arash Eshghi and Helen Hastie

*PGTask: Introducing the Task of Profile Generation from Dialogues*

Rui Ribeiro, Joao Paulo Carvalho and Luisa Coheur

*Question Generation to Elicit Users' Food Preferences Considering the Semantic Content*

Jie Zeng, Yukiko Nakano and Tatsuya Sakato

*Roll Up Your Sleeves: Working with a Collaborative and Engaging Task-Oriented Dialogue System*

Lingbo Mo, Shijie Chen, Ziru Chen, Xiang Deng, Ashley Lewis, Sunit Singh, Samuel Stevens, Chang-You Tai, Zhen Wang, Xiang Yue, Tianshu Zhang, Yu Su and Huan Sun

**15:10–15:40 Coffee Break**

**15:40–17:00 Oral Session 2: LLM for dialogue**

*Leveraging Large Language Models for Automated Dialogue Analysis*

Sarah E. Finch, Ellie S. Paek and Jinho D. Choi

*Are Large Language Models All You Need for Task-Oriented Dialogue?*

Vojtěch Hudeček and Ondrej Dusek

*Multi-party Goal Tracking with LLMs: Comparing Pre-training, Fine-tuning, and Prompt Engineering*

Angus Addlesee, Weronika Sieińska, Nancie Gunson, Daniel Hernandez Garcia, Christian Dondrup and Oliver Lemon

*ChatGPT vs. Crowdsourcing vs. Experts: Annotating Open-Domain Conversations with Speech Functions*

Lidiia Ostyakova, Veronika Smilga, Kseniia Petukhova, Maria Molchanova and Daniel Kornev

**Wednesday September 13, 2023 (continued)**

**17:10–18:10 Keynote: Aligning ChatGPT: past, present, and future**

**18:10–18:45 Panel Discussion: Social Impact of LLMs**

**19:00–20:00 Welcome Reception**

**Thursday September 14, 2023**

**09:00–10:00 Keynote: Engaging the Patient in Healthcare: Summarization and Interaction**

**10:00–10:30 Coffee Break**

**10:30–12:10 Oral Session 3: Dialogue modeling and evaluation**

*DiactTOD: Learning Generalizable Latent Dialogue Acts for Controllable Task-Oriented Dialogue Systems*

Qingyang Wu, James Gung, Raphael Shu and Yi Zhang

*Approximating Online Human Evaluation of Social Chatbots with Prompting*

Ekaterina Svikhnushina and Pearl Pu

*Dialogue Response Generation Using Completion of Omitted Predicate Arguments Based on Zero Anaphora Resolution*

Ayaka Ueyama and Yoshinobu Kano

*Syndicom: Improving Conversational Commonsense with Error-Injection and Natural Language Feedback*

Christopher Richardson and Larry Heck

*"What do others think?": Task-Oriented Conversational Modeling with Subjective Knowledge*

Chao Zhao, Spandana Gella, Seokhwan Kim, Di Jin, Devamanyu Hazarika, Alexandros Papangelis, Behnam Hedayatnia, Mahdi Namazifar, Yang Liu and Dilek Hakkani-Tur



**Thursday September 14, 2023 (continued)**

**12:10–13:00 SIGDIAL Business meeting**

**12:10–13:00 SIGGEN Business meeting**

**13:00–14:00 Lunch**

**14:00–15:40 Poster Session 2**

*UD\_Japanese-CEJC: Dependency Relation Annotation on Corpus of Everyday Japanese Conversation*

Mai Omura, Hiroshi Matsuda, Masayuki Asahara and Aya Wakasa

*Unravelling Indirect Answers to Wh-Questions: Corpus Construction, Analysis, and Generation*

Zulipiye Yusupujiang and Jonathan Ginzburg

*A New Dataset for Causality Identification in Argumentative Texts*

Khalid Al Khatib, Michael Voelske, Anh Le, Shahbaz Syed, Martin Potthast and Benno Stein

*Controllable Generation of Dialogue Acts for Dialogue Systems via Few-Shot Response Generation and Ranking*

Angela Ramirez, Kartik Agarwal, Juraj Juraska, Utkarsh Garg and Marilyn Walker

*Reference Resolution and New Entities in Exploratory Data Visualization: From Controlled to Unconstrained Interactions with a Conversational Assistant*

Abari Bhattacharya, Abhinav Kumar, Barbara Di Eugenio, Roderick Tabalba, Jillian Aurisano, Veronica Grosso, Andrew Johnson, Jason Leigh and Moira Zellner

*CONVERSER: Few-shot Conversational Dense Retrieval with Synthetic Data Generation*

Chao-Wei Huang, Chen-Yu Hsu, Tsu-Yuan Hsu, Chen-An Li and Yun-Nung Chen

*Speaker Role Identification in Call Centre Dialogues: Leveraging Opening Sentences and Large Language Models*

Minh-Quoc Nghiem, Nichola Roberts and Dmitry Sityaev

*Synthesising Personality with Neural Speech Synthesis*

Shilin Gao, Matthew P. Aylett, David A. Braude and Catherine Lai

**Thursday September 14, 2023 (continued)**

*Prompting, Retrieval, Training: An exploration of different approaches for task-oriented dialogue generation*

Gonçalo Raposo, Luisa Coheur and Bruno Martins

*Bootstrapping a Conversational Guide for Colonoscopy Prep*

Pulkit Arya, Madeleine Bloomquist, SUBHANKAR CHAKRABORTY, Andrew Perrault, William Schuler, Eric Fosler-Lussier and Michael White

*Applying Item Response Theory to Task-oriented Dialogue Systems for Accurately Determining User's Task Success Ability*

Ryu Hirai, Ao Guo and Ryuichiro Higashinaka

*An Open-Domain Avatar Chatbot by Exploiting a Large Language Model*

Takato Yamazaki, Tomoya Mizumoto, Katsumasa Yoshikawa, Masaya Ohagi, Toshiki Kawamoto and Toshinori Sato

**15:40–16:10 Coffee Break**

**16:10–17:50 Oral Session 3: Language understanding and multimodality**

*Learning Multimodal Cues of Children's Uncertainty*

Qi Cheng, Mert Inan, Rahma Mbarki, Grace Grmek, Theresa Choi, Yiming Sun, Kimele Persaud, Jenny Wang and Malihe Alikhani

*Grounding Description-Driven Dialogue State Trackers with Knowledge-Seeking Turns*

Alexandru Coca, Bo-Hsiang Tseng, Jinghong Chen, Weizhe Lin, Weixuan Zhang, Tisha Anders and Bill Byrne

*Resolving References in Visually-Grounded Dialogue via Text Generation*

Bram Willemsen, Livia Qian and Gabriel Skantze

*Slot Induction via Pre-trained Language Model Probing and Multi-level Contrastive Learning*

Hoang Nguyen, Chenwei Zhang, Ye Liu and Philip Yu

*The timing bottleneck: Why timing and overlap are mission-critical for conversational user interfaces, speech recognition and dialogue systems*

Andreas Liesenfeld, Alianda Lopez and Mark Dingemans

**Thursday September 14, 2023 (continued)**

**17:35–18:30 GenChal Poster Session + demos**

**19:00–22:00 Conference Dinner**

**Friday September 15, 2023**

**09:00–10:00 Virtual Poster Session**

*Enhancing Task Bot Engagement with Synthesized Open-Domain Dialog*

Miaoran Li, Baolin Peng, Michel Galley, Jianfeng Gao and Zhu (Drew) Zhang

*Enhancing Performance on Seen and Unseen Dialogue Scenarios using Retrieval-Augmented End-to-End Task-Oriented System*

Jianguo Zhang, Stephen Roller, Kun Qian, Zhiwei Liu, Rui Meng, Shelby Heinecke, Huan Wang, silvio savarese and Caiming Xiong

*Transformer-based Multi-Party Conversation Generation using Dialogue Discourse Acts Planning*

Alexander Chernyavskiy and Dmitry Ilvovsky

*Incorporating Annotator Uncertainty into Representations of Discourse Relations*

S. Magalí López Cortez and Cassandra L. Jacobs

*Investigating the Representation of Open Domain Dialogue Context for Transformer Models*

Vishakh Padmakumar, Behnam Hedayatnia, Di Jin, Patrick Lange, Seokhwan Kim, Nanyun Peng, Yang Liu and Dilek Hakkani-Tur

*C3: Compositional Counterfactual Contrastive Learning for Video-grounded Dialogues*

Hung Le, Nancy Chen and Steven C.H. Hoi

*No that's not what I meant: Handling Third Position Repair in Conversational Question Answering*

Vevake Balaraman, Arash Eshghi, Ioannis Konstas and Ioannis Papaioannou

*When to generate hedges in peer-tutoring interactions*

Alafate Abulimiti, Chloé Clavel and Justine Cassell

*PaperPersiChat: Scientific Paper Discussion Chatbot using Transformers and Discourse Flow Management*

Alexander Chernyavskiy, Max Bregeda and Maria Nikiforova

**Friday September 15, 2023 (continued)**

*FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions*

Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afil Farhan, Birthe Nessel, Meriam Moujahid, Tanvi Dinkar, Verena Rieser and Oliver Lemon

**10:00–11:00 Keynote: Knowledge graph use cases in natural language generation**

**11:00–11:30 Coffee Break**

**11:30–13:10 Oral Session 5: Topics in open-domain dialogue**

*Towards Breaking the Self-imposed Filter Bubble in Argumentative Dialogues*

Annalena Aicher, Daniel Kornmueller, Yuki Matsuda, Stefan Ultes, Wolfgang Minker and Keiichi Yasumoto

*The Open-domain Paradox for Chatbots: Common Ground as the Basis for Human-like Dialogue*

Gabriel Skantze and A. Seza Doğruöz

*MERCY: Multiple Response Ranking Concurrently in Realistic Open-Domain Conversational Systems*

Sarik Ghazarian, Behnam Hedayatnia, Di Jin, Sijia Liu, Nanyun Peng, Yang Liu and Dilek Hakkani-Tur

*Empathetic Response Generation for Distress Support*

Anuradha Welivita, Chun-Hung Yeh and Pearl Pu

*Reasoning before Responding: Integrating Commonsense-based Causality Explanation for Empathetic Response Generation*

Yahui Fu, Koji Inoue, Chenhui Chu and Tatsuya Kawahara

**Friday September 15, 2023 (continued)**

**13:10–14:10 Lunch**

**14:10–14:40 Sponsors**

**14:40–15:00 Closing**

**15:00–15:30 Birds-of-Feather**

**15:00–15:30 Coffee Break**

## Keynote Abstracts

### **Keynote 1: Engaging the Patient in Healthcare: Summarization and Interaction**

*Barbara Di Eugenio*

*University of Illinois, Chicago*

**Abstract:** Effective and compassionate communication with patients is becoming central to healthcare. I will discuss the results of and lessons learned from three ongoing projects in this space. The first, MyPHA, aims to provide patients with a clear and understandable summary of their hospital stay, which is informed by doctors' and nurses' perspectives, and by the strengths and concerns of the patients themselves. The second, SMART-SMS, models health coaching interactions via text exchanges that encourage patients to adopt specific and realistic physical activity goals. The third, HFChat, envisions an always-on-call conversational assistant for heart failure patients, that they can ask for information about lifestyle issues such as food and exercise. All our work is characterized by: large interdisciplinary groups of investigators who bring different perspectives to the research; grounding computational models in ecologically valid data, which is small by its own nature; the need for culturally valid interventions, since our UI Health system predominantly serves underprivileged, minority populations; and the challenges that arise when dealing with the healthcare enterprise.

**Bio:** Barbara Di Eugenio is a Professor and Director of Graduate Studies in the Computer Science department at the University of Illinois Chicago. There she leads the NLP laboratory (<http://nlp.cs.uic.edu/>). She obtained her PhD in Computer Science from the University of Pennsylvania (1993). Her research has always focused on the pragmatics and computational modeling of discourse and dialogue, grounded in authentic data collection on the one hand, and in user studies on the other. The applications of her work run the gamut from educational technology to human-robot interaction, from data visualization to health care. Dr. Di Eugenio is an NSF CAREER awardee (2002); a UIC University Scholar (2018-2020); and a Zenith Award recipient from AWIS, the Association for Women in Science (2022). She has been the editor-in-chief for the Journal of Discourse and Dialogue since 2019. She is very proud to have graduated 15 PhD and 32 Master's students.

## **Keynote 2: Textless NLP: towards language processing from raw audio**

*Emmanuel Dupoux*

*Ecole des Hautes Etudes en Sciences Sociales (EHESS)*

**Abstract:** The oral (or gestural) modality is the most natural channel for human language interactions. Yet, language technology (Natural Language Processing, NLP) is primarily based on the written modality, and requires massive amounts of textual resources for the training of useful language models. As a result, even fundamentally speech-first applications like speech-to-speech translation or spoken assistants like Alexa, or Siri, are constructed in a Frankenstein way, with text as an intermediate representation between the signal and language models. Besides this being inefficient, This has two unfortunate consequences: first, only a small fraction of the world’s languages that have massive textual repositories can be addressed by current technology. Second, even for text-rich languages, the oral form mismatches the written form at a variety of levels, including vocabulary and expressions. The oral medium also contains typically unwritten linguistic features like rhythm and intonation (prosody) and rich paralinguistic information (non verbal vocalizations like laughter, cries, clicks, etc, nuances carried through changes in voice qualities) which are therefore inaccessible to language models. But is this a necessity? Could we build language applications directly from the audio stream without using any text? In this talk, we review recent breakthroughs in representation learning and self-supervised techniques which have made it possible to learn latent linguistic units directly from audio which unlock the learning of generative language models without the use of any text. We show that these models can capture heretofore un-addressed nuances of the oral language including in a dialogue context, opening up the possibility of speech-to-speech textless NLP applications. We outline existing technical challenges to achieve this goal, including challenges to build expressive oral language datasets at scale.

**Bio:** E. Dupoux is professor at the Ecole des Hautes Etudes en Sciences Sociales (EHESS) and Research Scientist at Meta AI Labs. He directs the Cognitive Machine Learning team at the Ecole Normale Supérieure (ENS) in Paris and INRIA. His education includes a PhD in Cognitive Science (EHESS), a MA in Computer Science (Orsay University) and a BA in Applied Mathematics (Pierre & Marie Curie University). His research mixes developmental science, cognitive neuroscience, and machine learning, with a focus on the reverse engineering of infant language and cognitive development using unsupervised or weakly supervised learning. He is the recipient of an Advanced ERC grant, co-organizer of the Zero Ressource Speech Challenge series (2015–2021), the Intuitive Physics Benchmark (2019) and led in 2017 a Jelinek Summer Workshop at CMU on multimodal speech learning. He is a CIFAR LMB and a ELLIS Fellow. He has authored 150 articles in peer reviewed outlets in cognitive science and language technology.

### **Keynote 3: Knowledge graph use cases in natural language generation**

*Elena Simperl*

*King's College London*

**Abstract:** Natural language generation (NLG) makes knowledge graphs (KGs) more accessible. I will present two applications of NLG in this space: in the first one, verbalisations of KG triples feed into downstream KG applications, allowing users with diverse levels of digital literacy to share their knowledge, and contribute to the KG. In the second one, having text representations of KG triples helps us verify the content of a KG against external sources towards more trustworthy KGs. I will present human-in-the-loop solutions to these applications that leverage a range of machine learning techniques to scale to the large, multilingual knowledge graphs modern applications use.

**Bio:** Elena Simperl is a Professor of Computer Science and Deputy Head of Department for Enterprise and Engagement in the Department of Informatics at King's College London. She is also the Director of Research for the Open Data Institute (ODI) and a Fellow of the British Computer Society and the Royal Society of Arts. Elena features in the top 100 most influential scholars in knowledge engineering of the last decade. She obtained her doctoral degree in Computer Science from the Free University of Berlin, and her diploma in Computer Science from the Technical University of Munich. Prior to joining King's in 2020, she was a Turing Fellow, and held positions in Germany, Austria and at the University of Southampton. Her research is at intersection between AI and social computing, helping designers understand how to build smart sociotechnical systems that combine data and algorithms with human and social capabilities. Elena led 14 European and national research projects, including recently QROWD, ODINE, Data Pitch, Data Stories, and ACTION. She is currently the scientific and technical director of MediaFutures, a Horizon 2020 programme that is using arts-inspired methods to design participatory AI systems that tackle misinformation and disinformation online. Elena's interest in leading initiatives within the scientific community has also taken form through chairing several conferences in her field, including the European and International Semantic Web Conference series, the European Data Forum, and the European Semantic Technologies conference. She is the president of the Semantic Web Science Association.



#### **Keynote 4: Aligning ChatGPT: past, present, and future**

*Ryan Lowe*

*OpenAI*

**Abstract:** In this talk I will present different perspectives on the alignment of chatbots like ChatGPT. I'll review reinforcement learning from human feedback (RLHF), the core training technique behind InstructGPT and ChatGPT, including a brief history of how it was developed. I'll discuss some of the pitfalls of RLHF, and what is being done today to address them. I'll then speculate on some of the alignment challenges I expect we'll face with this new generation of powerful personal assistants, how they could reshape society, and some things we'll need to do to make sure these changes are good for humans.

**Bio:** Ryan is a researcher at OpenAI on the Alignment team. His most recent work involved proving out RLHF on language models, starting with summarization, then moving to InstructGPT and most recently ChatGPT and GPT-4. Previously, he worked on multi-agent RL, emergent communication, and dialogue systems at McGill University.

