YRRSDS 2023

The 19th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems

Proceedings of the Workshop

September 11 - 12, 2023
Prague, Czechia

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 978-1-952148-25-5

Sponsor

In Collaboration With

Preface

We are delighted to provide the opening words for the proceedings of the 19th Young Researchers Roundtable on Spoken Dialogue Systems (YRRSDS) 2023, a workshop on Spoken Dialogue Systems for PhDs, PostDocs and New Researchers. YRRSDS 2023 was collocated with the Special Interest Group on Discourse and Dialogue (SIGDIAL) 2023. The workshop took place on September 11-12, 2023 in Prague, Czech Republic at the OREA Hotel Pyramida. This year, the format for YRRSDS was in person.

Submissions to YRRSDS consisted of writing a 2 page position paper outlining the young researcher's current research topic and interests, and general points they would like to see discussed during the roundtable sessions in the workshop.

Each submission was reviewed by 2 senior researchers from our Advisory Committee. We are immensely grateful to the members of the Advisory Committee for their excellent and thoughtful reviews. Their contributions have been essential to providing critical feedback to the participants of the workshop during this stage in their career.

Participants accepted to the program were required to present a poster based on their submission. This year, YRRSDS accepted all 25 submissions that were received.

The roundtable discussions this year focused on the following topics: Large Language Models (LLMs), Evaluation methods in Spoken Dialogue Systems, Knowledge Bases, Reasoning & Planning, Multimodality & Interaction, Architectures and Ethics, Privacy & Regulations.

In addition to the poster sessions and roundtables, the program consisted of 3 fantastic keynote presentations. We would like to take this opportunity to thank and acknowledge our 3 keynote speakers: Verena Rieser (Senior Staff Research Scientist at Google DeepMind, hon. Professor at Heriot-Watt University and Co-Founder at ALANA AI), Malihe Alikhani (Assistant Professor of AI and social justice at the Khoury School of Computer Science, Northeastern University) and David Traum (Director for Natural Language Research at the Institute for Creative Technologies and Research Professor at USC Viterbi School of Engineering Computer Science Department) for their inspiring talks.

We thank the organisers who ensured that the conference ran very smoothly, and was enjoyed by all participants. We gratefully acknowledge the support of our sponsor: Omilia Conversational Intelligence.

Lastly, we are excited to announce that this year will mark the first year the YRRSDS proceedings will be integrated as part of the ACL anthology. We hope that by integrating the proceedings in this way, young researchers will have the opportunity to have visibility on their positional submission, and by extension their research.

Tanvi Dinkar and Javier Chiyah-Garcia, Organizing Committee YRRSDS 2023

Organizing Committee

Organizers:

Vojtěch Hudeček

Charles University

Patrícia Schmidtová

Charles University

Tanvi Dinkar

Heriot-Watt University

Weronika Sieińska

Heriot-Watt University

Javier Chiyah-Garcia

Heriot-Watt University

Advisory committee:

Srinivas Bangalore, *Interactions LLC*

Timo Baumann, *University of Hamburg, OTH Regensburg*

Luciana Benotti, *National University of Córdoba*

Hendrik Buschmeier, *Bielefeld University*

Justine Cassell, *PRAIRIE Paris Artificial Intelligence Research Institute in INRIA Paris*

Nina Dethlefs, *University of Hull*

Luis Fernando D'Haro Enríquez, *Technical University of Madrid*

Kallirroi Georgila, *University of Southern California*

Larry Heck, *Viv Labs*

Ryuichiro Higashinaka, *Nagoya University*

Julia Hirschberg, *Columbia University*

Krzysztof Jassem, *Adam Mickiewicz University*

Tatsuya Kawahara, *Kyoto University*

James Kennedy, *Disney Research*

Kazunori Komatani, *Osaka University*

Ioannis Konstas, *Heriot-Watt University*

Udo Kruschwitz, *Regensburg University*

Marek Kubis, *Adam Mickiewicz University*

Jackson Liscombe, *Modality AI*

Pierre Lison, *Norwegian Computing Centre*

Wolfgang Minker, *University of Ulm*

Sebastian Möller, *Technical University of Berlin*

Mikio Nakano, *C4A Research Institute*

Diarmuid Ó Séaghdha, *Apple*

Alexandros Papangelis, *Amazon Alexa AI*

Vikram Ramanarayanan, *Modality AI*

Abhinav Rastogi, *Google*

Giuseppe Riccardi, *University of Trento*

David Schlangen, *University of Potsdam*

Gabriel Skantze, *KTH Royal Institute of Technology*

Paweł Skórzewski, *Adam Mickiewicz University*

David Traum, *University of Southern California*

Khiet Truong, *University of Twente*

Stefan Ultes, *University of Bamberg*

Marilyn Walker, *University of California, Santa Cruz*

Nigel Ward, *University of Texas at El Paso*

Michael White, *Ohio State University*

Zhou Yu, *Columbia University*

Organizers' Notes of the Roundtable Discussions

Ethics, Privacy & Regulations

The meeting's primary focus revolved around the ethical aspects of privacy, regulation, and human interaction with virtual systems. The conversation began with a deep dive into privacy, highlighting the delicate balance between privacy and utility. Participants stressed the importance of making informed choices when it comes to cookies and being aware of how data is used, especially with virtual systems that can blur the lines. An interesting point was raised about data usage by companies, as some have recently updated their policies regarding collecting user data based on input queries into LLMs.

Anthropomorphism in dialogue systems emerged as another important topic. The use of human-like language to describe these systems, along with customizable chatbots, raised questions about how users perceive these systems and the impact on the community. The need for clarifying the relationship between users and the system was emphasised, with user studies to explore confirmation bias when integrating users into the dialogue process.

Architecture

The meeting involved a discussion on end-to-end (E2E) systems versus modular systems in dialogue systems, with E2E systems increasingly becoming the standard while modular systems are less favoured. However, E2E systems have issues related to low interpretability. The question of whether Language Models (LLMs) are better suited than modular architectures for production environments was raised, with the consensus that LLMs tend to work efficiently, even with small amounts of data. The discussion also touched on the trade-off between the size and speed of integrating LLMs in a dialogue system. In the context of transformers, memory constraints and limitations were examined, highlighting their short memory, which can be a limitation in dialogue settings unless trained with large sequence lengths.

Large Language Models

During the meeting, the attendees discussed the suitability of Language Models (LLMs) specifically applied to Spoken Dialogue Systems (SDS). Some researchers discussed their experiences with multi-party dialogues, where LLMs surprisingly demonstrated effectiveness in various conversational scenarios. This raised questions about the potential practical applications of LLMs in complex dialogue systems.

However, the real-world deployment of LLMs raised concerns. One significant issue was hallucinations, where LLMs generate information that is incorrect or entirely fictional. These hallucinations can be problematic and affect the reliability of the model's output. Also discussed were the consequences of training LLMs with synthetic data generated by other LLMs, emphasising potential complications and unintended outcomes. The discussion shifted to the uncertainty surrounding the data used to train LLMs, as well as the risk of data contamination. It remains unclear what information these models have been exposed to during training. Another concerning aspect was the overconfidence of LLMs in their answers. This underscores the importance of fact-checking and ensuring the reliability of information provided by these models.

Despite these challenges, the meeting recognized the positive impacts of LLMs. They can be valuable in tasks like enhancing the fluency and functionality of dialogue systems. High-quality English LLMs can assist in language learning and provide access to knowledge. Furthermore, LLMs have demonstrated strengths in picking up on nuances in task-oriented dialogues.

Finally, the conversation touched on fundamental limitations in LLM architecture. Predicting the next word was considered a limitation, especially when dealing with long-context dialogues and multi-modal tasks.These limitations were central to the debate surrounding the overall capabilities and potential constraints of LLMs in the field of dialogue and communication.

Evaluation

The discussion started with a focus on the quality of evaluations conducted by crowd workers and the inherent challenges when utilising Mechanical Turk for human evaluations in Natural Language Processing (NLP). Anecdotes about experiences with Mechanical Turk were shared, shedding light on issues related to privacy and high rejection rates. The challenges of managing rejection rates and compensation for incomplete tasks on Mechanical Turk were explored in detail.

As the dialogue progressed, various crowd-sourcing platforms, including alternatives like Prolific, were brought into consideration for specific survey needs and participant selection. This shift highlighted the need for identifying and filtering out low-quality annotators, with attention checks and quality control measures being proposed as potential solutions. The conversation also delved into the reproducibility of human evaluations, even for apparently objective tasks such as fluency and grammaticality.

Lastly, the granularity of annotations for dialogue research was discussed; i.e. shifting from collecting annotations at the system-level to the turn-level. The conversation also discussed the idea of collecting diverse opinions from annotators and employing distributional approaches to assess model performance.

Knowledge Bases, Reasoning & Planning

The discussion centred around how to create solutions to dynamically update or access knowledge within LLMs that have been integrated into spoken dialogue systems. Since information is constantly evolving, having a static knowledge base can lead to outdated and inaccurate responses in a dialogue system.

We also discussed whether querying LLMs with specific knowledge requests or adopting more complex integration methods would be more effective for dialogue systems. The choice between these approaches is pivotal as it impacts the user experience and the reliability of information provided.

However, a significant concern that emerged during our discussion was that LLMs produce hallucinations and generate responses that are factually incorrect or misleading.

Multi-Modality & Interaction

The discussion was centred around the nature of multi-modal conversations, which is very subjective. The limitations of using only text was emphasised, especially in capturing the contextual cues of dialogue, such as speech, tone, and prosody, highlighting how multimodal cues have the potential to alter conversation meanings.

It was noted that many works claim multi-modality, even when they predominantly involve image and text, indicating that the term has become somewhat overused.

A question posed was whether multi-modality could provide the missing context in current spoken dialogue systems. Examples of multi-modality in spoken dialogue systems were then shared. It was stressed that integrating multi-modality into SDS is crucial, but significantly more challenging compared to non-interactive models.

Table of Contents

Conference Program

Monday, September 11, 2023

09:30–09:45 Registration and Poster Setup

09:45–10:00 Welcome

10:00–10:30 Industry Keynote: Omilia (Gold Sponsor) by Vojtěch Hudeček

10:30–11:00 Coffee Break

11:00–12:00 Poster Session 1

Processing Referential Ambiguities in Situated Dialogue Systems
Javier Chiyah-Garcia

Safety and Robustness in Conversational AI
Tanvi Dinkar

Incremental Speech Processing for Voice Assistant Accessibility
Angus Addlesee

Advancing Spoken Dialog Systems for Manufacturing: From Conceptual Architecture and Taxonomy to Real Case Applications and Future Directions
Silvia Colabianchi

Conversational Grounding in Multimodal Dialog Systems
Biswesh Mohapatra

SQL Comment Generation and Additional Research Interests
Alyssa Allen

On Referring Language Use in Visually Grounded Dialogue
Bram Willemsen

Challenges and Approaches in Designing Social SDS in the LLM Era
Koji Inoue

Breakdowns and Repairs. Detecting Patterns that Lead to Breakdowns in Customer Service Messages
Anouck Braggaar

Towards More Natural Dialogues: Integrating Open-Domain Dialogue Skills into Task-Oriented Agents
Armand Stricker

The Future of Designing Spoken Dialogue Systems and Analyzing Written Conversations
Livia Qian

Exploring the Synergy of Deep Learning and Anthropomorphism in Multimodal Dialogue Systems
Iwona Christop

12:00–12:30 Roundtable Session 1: Ethics, Privacy & Regulations
Chairs: Tanvi Dinkar, Patrícia Schmidtová

12:00–12:30 Roundtable Session 1: Architecture
Chair: Javier Chiyah-Garcia

12:30–13:30 Lunch

13:30–14:30 Poster Session 2

A Perspective on Anchoring and Dialogue History Propagation for Smoother Interactions with Spoken Task-Oriented Dialogue Systems
Lucas Druart

More Human-Like Interaction in Spoken Dialogue Systems: Global Context for Natural Language Understanding and Multimodal Solutions
Kacper Dudzic

Designing and Evaluating LLM-based Conversational Agents for Behaviour Change
Selina Meyer

Monday, September 11, 2023 (continued)

14:30–15:15 **Keynote: A short history of data-driven dialogue systems in 5 acts: Where do we go from here? by Verena Rieser**

15:15–15:45 **Coffee Break**

15:45–16:30 **Roundtable Session 2: Large Language Models**

19:30 **Dinner**

Tuesday, September 12, 2023

09:30–09:45 **Welcome**

09:45–10:30 **Roundtable Session 3: Evaluation**
Chairs: Patrícia Schmidtová, Tanvi Dinkar

09:45–10:30 **Roundtable Session 3: Emotion, Empathy & Personalised Dialogues**
Chair: Vojtěch Hudeček

10:30–11:00 **Coffee Break**

11:00–11:45 **Roundtable Session 4: Knowledge Bases, Reasoning & Planning**
Chair: Vojtěch Hudeček

Tuesday, September 12, 2023 (continued)

11:00–11:45 **Roundtable Session 4: Multi-Modality & Interaction**
Chair: Javier Chiyah-Garcia

11:45–12:30 **Keynote: The past, present, and future of dialogue systems and advice to young researchers by David Traum**

12:30–13:30 **Lunch**

13:30–14:00 **Group Photo**

14:00–14:45 **Keynote: Leveraging Generative AI for Inclusive and Equitable Dialogue Systems by Malihe Alikhani**

14:45–15:00 **Closing**

15:00 **Free Discussion**

Keynotes

Keynote 1: A short history of data-driven dialogue systems in 5 acts: Where do we go from here?

Verena Rieser

Senior Staff Research Scientist at Google DeepMind, hon. Professor at Heriot-Watt University and Co-Founder at ALANA AI

Bio: Verena is a Senior Staff Research Scientist at Google DeepMind, where she works on Safer Conversational AI. She is also honorary professor at Heriot-Watt University in Edinburgh and a co-founder of the Conversational AI company ALANA AI. Verena holds a PhD from Saarland University in Germany and a MSc from the University of Edinburgh, where she also spent time as a postdoctoral researcher.

She has 20 years of experience in developing and researching data-driven conversational systems. In the early 2000s she developed a series of breakthrough innovations that laid the groundwork for statistical dialogue control using Reinforcement Learning. More recently, Verena and her team pioneered work on identifying and addressing safety risks in neural conversational systems, which was awarded with a Leverhulme Senior Research Fellowship by the Royal Society.

Keynote 2: The past, present, and future of dialogue systems and advice to young researchers

David Traum

Director for Natural Language Research at the Institute for Creative Technologies (ICT) and Research Professor at USC Viterbi School of Engineering Computer Science Department

Bio: David Traum is a principal scientist at ICT and a research faculty member at the Department of Computer Science at USC. At ICT, Traum leads the Natural Language Dialogue Group, which consists of seven Ph.D.s, four students, and four other researchers.

The group engages in research in all aspects of natural language dialogue, including dialogue management, spoken and natural language understanding and generation and dialogue evaluation. In addition, the group collaborates with others at ICT and elsewhere on integrated virtual humans, and transitioning natural language dialogue capability for use in training and other interactive applications.

Traum's research focuses on dialogue communication between human and artificial agents. He has engaged in theoretical, implementational and empirical approaches to the problem, studying human-human natural language and multi-modal dialogue, as well as building a number of dialogue systems to communicate with human users.

He has pioneered several research thrusts in computational dialogue modeling, including computational models of grounding (how common ground is established through conversation), the information state approach to dialogue, multiparty dialogue, and non-cooperative dialogue.

Traum is author of over 200 technical articles, is a founding editor of the Journal Dialogue and Discourse, has chaired and served on many conference program committees, and is currently the president emeritus of SIGDIAL, the international special interest group in discourse and dialogue. He earned his Ph.D. in computer science at University of Rochester in 1994.

Keynote 3: Leveraging Generative AI for Inclusive and Equitable Dialogue Systems

Malihe Alikhani

Assistant Professor of AI and social justice at the Khoury School of Computer Science, Northeastern University

Bio: Malihe Alikhani is an Assistant Professor at the Khoury School of Computer Science, Northeastern University. She earned her Ph.D. in computer science with a graduate certificate in cognitive science from Rutgers University in 2020. Her research interests center on using representations of communicative structure, machine learning, and cognitive science to design equitable and inclusive language technologies. This involves developing systems that can communicate and collaborate with diverse populations, especially those from underserved communities, for critical applications such as education, health, and social justice. Her work has received best paper awards at ACL 2021, UAI2022, INLG2021, and UMAP2022 and has been supported by DARPA, NIH, Google, and Amazon.