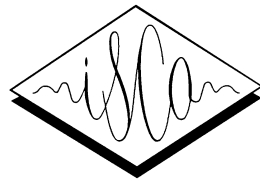


SIGDIAL 2021

**22nd Annual Meeting of the  
Special Interest Group on Discourse and Dialogue**



**Proceedings of the Conference**

29-31 July 2021  
Singapore and Online

**In cooperation with:**

Association for Computational Linguistics (ACL)

International Speech Communication Association (ISCA)

Association for the Advancement of Artificial Intelligence (AAAI)

**We thank our sponsors:**

- LivePerson
- Apple
- DataBaker (Beijing) Technology
- Google
- Rasa Technologies
- Furhat Robotics
- Toshiba Research Europe
- Chinese and Oriental Languages Information Processing Society
- National University of Singapore
- Teochew Doctorate Society Singapore

**Platinum**



**Gold**



Silver



**TOSHIBA**

Local Sponsors

**COLIPS**



新加坡潮籍博士會  
*Teochew Doctorate Society • Singapore*

©2021 The Association for Computational Linguistics

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 978-1-954085-81-7

## Preface

We are glad to pen the first few words for the proceedings of SIGDIAL 2021, the 22nd 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 as a hybrid event with both in-person and virtual participation on July 29-31, 2021, right before ACL-IJCNLP 2021. The 2021 Young Researchers' Roundtable on Spoken Dialog Systems (YRRSDS 2021) is also held as a satellite event. The SIGDIAL 2021 program features three keynote talks, 6 paper presentation sessions, 1 demo session, and 2 special sessions, entitled "Summarization of Dialogues and Multi-Party Meetings", and "Safety for E2E Conversational AI".

COVID has changed the way we work, but it doesn't hamper our research progress. We received 142 submissions this year, comprising 88 long papers, 49 short papers, and 5 demo descriptions. We had 12 Senior Program Committee (SPC) members who were each responsible for 11-12 papers, leading the discussion process and also contributing 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 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. We have aimed to develop a broad, varied program spanning the many positively rated papers identified by the review process. We accepted 59 papers in total: 40 long papers (45%), 15 short papers (31%), and 4 demo descriptions, for an overall acceptance rate of 41.5%, in line with prior years.

One keynote will highlight each of the three days of the conference. In organizing this hybrid in-person/virtual conference, we have tried to maintain as much of the spirit of a fully online conference as possible. Recordings for all papers and demos have been made available several days before the start of the conference, for participants to watch asynchronously. Long and short papers are organized into sessions taking into consideration the presenters' time zones. Regular papers sessions span 8-11 papers, each presented as a two-minute pre-recorded talk followed by five minutes of live Q&A. For demos, we organized four parallel zoom rooms to allow participants to interact with and observe live interactions with the systems. The topics represent the breadth of research in discourse and dialogue. A conference of this size 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 three keynote speakers, Julia Hirschberg (Columbia University), Raymond J. Mooney (University of Texas at Austin), and Jason Weston (Facebook AI & NYU), for their inspiring talks on "Whom Do We Trust in Dialogue Systems?", "Dialog with Robots: Perceptually Grounded Communication with Lifelong Learning", and "A journey from ML & NNs to NLP and Beyond: Just more of the same isn't enough?" We also thank the organizers of the two special sessions: "Summarization of Dialogues and Multi-Party Meetings", and "Safety for E2E Conversational AI". We are grateful for their coordination with the main conference.

SIGDIAL 2021 is made possible by the dedication and hard work of our community. We are indebted to many. The SIGDIAL track record of excellence continues this year. This would not have been possible without the advice and support of the SIGDIAL board, particularly Gabriel Skantze and Mikio Nakano for their guidance. Special mention must be made of the fact that, for the first time, we pilot a hybrid conference to facilitate the participation. This inevitably increases the workload for the organizers.

We take this opportunity to express our gratitude to the local chairs, Chitralekha Gupta, and Berrak Sisman for coordinating everything flawlessly, the local co-chairs, Yi Zhou, Mingyang Zhang, Grandee Lee, Rui Liu, Zongyang Du, Kun Zhou, and Chen Zhang for managing the virtual platform and local

matters professionally; the COLIPS council members Yan Wu, Minghui Dong, and Lei Wang for their tremendous support to the arrangement of venue and social programs. Special thanks go to local chair Siqui Cai for her tireless effort in managing the website with timely updates, and to local co-chair Bidisha Sharma for conference registration, last but not least, to Celine Cheong and Min Yuan for their administrative support. SIGDIAL 2021 would not have been possible without their extraordinary effort.

We would also like to thank the sponsorship chair David Vandyke, who has been our SIGDIAL ambassador to the industry year after year. He continued to bring to the conference an impressive panel of conference sponsors. We thank David for his dedicated effort. We gratefully acknowledge the support of our sponsors: LivePerson (Platinum), Apple, DataBaker, Google and Rasa Technologies (Gold) and Furhat Robotics, Toshiba Research Europe (Silver). In addition, we thank Jessy Li, the publication chair, Nina Dethlefs, the mentoring chair for their dedicated services.

Finally, it is our great pleasure to welcome you physically and virtually to the conference. We hope that you will have an enjoyable and productive time, and leave with fond memories of SIGDIAL 2021. With our best wishes for a successful conference!

Haizhou Li, General Chair

Gina-Anne Levow, Zhou Yu, Program Co-Chairs



**General Chair:**

Haizhou Li, National University of Singapore, Singapore

**Program Chairs:**

Gina-Anne Levow, University of Washington, USA

Zhou Yu, Columbia University, USA

**Local Arrangements Team:**

Chitralkha Gupta (Chair), National University of Singapore, Singapore

Berrak Sisman (Chair), Singapore University of Technology and Design, Singapore

Siqi Cai (Chair), National University of Singapore, Singapore

Bidisha Sharma (Co-Chair, events, registration), National University of Singapore, Singapore

Yan Zhang (Co-Chair, technical program), National University of Singapore, Singapore

Chen Zhang (Co-Chair, technical program), National University of Singapore, Singapore

Grandee Lee (co-chair, student volunteers), National University of Singapore, Singapore

Yan Wu (Co-Chair, social events), A\*STAR Institute for Infocomm Research, Singapore

Lei Wang (COLIPS Liaison), COLIPS, Singapore

Celine Cheong (Co-Chair, venue), National University of Singapore, Singapore

Min Yuan (Co-Chair, venue), National University of Singapore, Singapore

Kun Zhou (Co-Chair, artwork), National University of Singapore, Singapore

Yi Zhou (Co-Chair, technical support), National University of Singapore, Singapore

Rui Liu (Co-Chair, technical support), National University of Singapore, Singapore

Mingyang Zhang (Co-Chair, technical support), National University of Singapore, Singapore

Minghui Dong (COLIPS board), A\*STAR Institute for Infocomm Research, Singapore

**Sponsorship Chair:**

David Vandyke, Apple Inc., United Kingdom

**Mentoring Chair:**

Nina Dethlefs, University of Hull, United Kingdom

**Finance Chair:**

Yan Wu, A\*STAR Institute for Infocomm Research, Singapore

**Publication Chair:**

Junyi Jessy Li, The University of Texas at Austin, USA

**SIGdial Officers:**

President: Gabriel Skantze, KTH Royal Institute of Technology, Sweden

Vice President: Mikio Nakano, C4A Research Institute, Japan

Secretary: Vikram Ramanarayanan, Educational Testing Service (ETS) Research, USA

Treasurer: Ethan Selfridge, LivePerson, USA

President Emeritus: Jason Williams, Apple, USA

**Senior Program Committee:**

Asli Celikyilmaz, Facebook AI

Vivian Chen, National Taiwan University

Katherine Forbes-Riley

Milica Gasic, Heinrich Heine University Düsseldorf

Ryuichiro Higashinaka, Nagoya University/NTT Media Intelligence Labs  
Annie Louis, Google  
Mikio Nakano, C4A Research Institute  
Vincent Ng, University of Texas at Dallas  
Rebecca J. Passonneau, Penn State University  
Gabriel Skantze, KTH  
David Traum, University of Southern California  
Koichiro Yoshino, RIKEN



## **Program Committee:**

Sean Andrist, Microsoft Research, United States  
Masahiro Araki, Kyoto Institute of Technology, Japan  
Timo Baumann, Universität Hamburg, Germany  
Frederic Bechet, Aix Marseille Université - LIS/CNRS, France  
Steve Beet, Aculab plc, United Kingdom  
Jose Miguel Benedi, Universitat Politècnica de València, Spain  
Luciana Benotti, Universidad Nacional de Córdoba, Argentina  
Parminder Bhatia, Amazon, United States  
Nate Blaylock, Canary Speech, United States  
Hendrik Buschmeier, Bielefeld University, Germany  
Andrew Caines, University of Cambridge, United Kingdom  
Patricia Chaffey, USC, United States  
Senthil Chandramohan, Microsoft, United States  
Lin Chen, Head of AI, Cambia Health Solutions, United States  
Derek Chen, ASAPP, United States  
Paul Crook, Facebook, United States  
Orianna Demasi, UC Davis, United States  
Nina Dethlefs, University of Hull, United Kingdom  
David DeVault, Anticipant Speech, Inc., United States  
Emily Dinan, Facebook AI Research, United States  
Maxine Eskenazi, Carnegie Mellon University, United States  
Mauro Falcone, Fondazione Ugo Bordoni, Italy  
Kallirroi Georgila, University of Southern California Institute for Creative Technologies, United States  
Felix Gervits, US Army Research Laboratory, United States  
David Gros, University of California - Davis, United States  
Jing Gu, University of California, Davis, United States  
Joakim Gustafson, KTH, Sweden  
Ivan Habernal, Technische Universität Darmstadt, Germany  
Dilek Hakkani-Tur, Amazon Alexa AI, United States  
Helen Hastie, Heriot-Watt University, United Kingdom  
Michael Heck, Heinrich Heine University, Germany  
Behnam Hedayatnia, Amazon, United States  
Peter Heeman, OHSU / CSLU, United States  
Takuya Hiraoka, NEC Central Research Laboratories, Japan  
Michimasa Inaba, The University of Electro-Communications, Japan  
Koji Inoue, Kyoto University, Japan  
Simon Keizer, Toshiba Europe Ltd, United Kingdom  
Casey Kennington, Boise State University, United States  
Kazunori Komatani, Osaka University, Japan  
Jared Kramer, Amazon, United States  
Ivana Kruijff-Korbayova, DFKI, Germany  
Kornel Laskowski, Carnegie Mellon University, United States  
Fabrice Lefèvre, Avignon Univ., France  
Kai-Hui Liang, Columbia University, United States  
Pierre Lison, Norwegian Computing Centre, Norway  
Bing Liu, Facebook, United States  
Eduardo Lleida Solano, University of Zaragoza, Spain  
José Lopes, Heriot Watt University, United Kingdom  
Nurul Lubis, Heinrich Heine University, Germany

Eleonore Lumer, Bielefeld University, Germany  
Ramesh Manuvinakurike, Intel labs, United States  
Teruhisa Misu, Honda Research Institute USA, United States  
Seungwhan Moon, Facebook Reality Labs, United States  
Elena Musi, University of Liverpool, United Kingdom  
Satoshi Nakamura, Nara Institute of Science and Technology and RIKEN AIP Center, Japan  
Anna Nedoluzhko, Charles University in Prague, Czech Republic  
Douglas O'Shaughnessy, INRS-EMT (Univ. of Quebec), Canada  
Alexandros Papangelis, Amazon Alexa AI, United States  
Aasish Pappu, Spotify Research, United States  
Paul Piwek, The Open University, United Kingdom  
Heather Pon-Barry, Mount Holyoke College, United States  
Shrimai Prabhumoye, Carnegie Mellon University, United States  
Stephen Pulman, Apple Inc., United Kingdom  
Kun Qian, Columbia University, United States  
Liang Qiu, University of California, Los Angeles, United States  
Vikram Ramanarayanan, University of California, San Francisco, United States  
Hannah Rashkin, Google Research, United States  
Verena Rieser, Heriot-Watt University, United Kingdom  
Antonio Roque, Tufts University, United States  
Carolyn Rosé, Carnegie Mellon University, United States  
Clayton Rothwell, Infocitex Corp., United States  
alexander rudnicky, Carnegie Mellon University, United States  
Saurav Sahay, Intel Labs, United States  
Sakriani Sakti, Nara Institute of Science and Technology (NAIST) / RIKEN AIP, Japan  
Chinnadhurai Sankar, Facebook AI, United States  
Ruhi Sarikaya, Amazon, United States  
Matthias Scheutz, Tufts University, United States  
Ethan Selfridge, LivePerson, United States  
Matthias Scheutz, Tufts University, United States  
Ethan Selfridge, LivePerson, United States  
Weiyan Shi, Columbia University, United States  
Aaron Sisto, Searchable.ai, United States  
Georg Stemmer, Intel Corp., Germany  
Svetlana Stoyanchev, Toshiba Europe, United Kingdom  
Kristina Striegnitz, Union College, United States  
Hiroaki Sugiyama, NTT Communication Science Labs., Japan  
António Teixeira, DETI/IEETA, University of Aveiro, Portugal  
Takenobu Tokunaga, Tokyo Institute of Technology, Japan  
Bo-Hsiang Tseng, University of Cambridge, United Kingdom  
Gokhan Tur, Amazon Alexa AI, United States  
Stefan Ultes, Mercedes-Benz AG, Germany  
David Vandyke, Apple, United Kingdom  
Yi-Chia Wang, Facebook AI, United States  
Nigel Ward, University of Texas at El Paso, United States  
Qingyang Wu, Columbia University, United States  
Koichiro Yoshino, RIKEN Robotics, Nara Institute of Science and Technology, Japan  
Steve Young, Cambridge University, United Kingdom  
Maryam Zare, Pennsylvania State University, United States  
Jian ZHANG, School of Computer Science, Dongguan University of Technology; Department of Computer Science and Engineering, the Hong Kong University of Science and Technology, China

Tiancheng Zhao, SOCO.AI, United States  
Mingyang Zhou, Ph.D. Student at University of California, Davis, United States  
Ingrid Zukerman, Monash University, Australia

**Secondary Reviewers:** Nicholas Walker, Dmytro Kalpakchi, Christian Geishauser, Tatiana Anikina,  
Wolfgang Maier, Ye Liu, Eda Okur, Mauricio Mazuecos

**Invited Speakers:**

Julia Hirschberg, Columbia University, USA  
Raymond J. Mooney, University of Texas, USA  
Jason Weston, Facebook AI & NYU Visiting Research Professor, USA



## Table of Contents

<i>Understanding and predicting user dissatisfaction in a neural generative chatbot</i> Abigail See and Christopher Manning .....	1
<i>Towards Continuous Estimation of Dissatisfaction in Spoken Dialog</i> Nigel Ward, Jonathan E. Avila and Aaron M. Alarcon .....	13
<i>DialogStitch: Synthetic Deeper and Multi-Context Task-Oriented Dialogs</i> Satwik Kottur, Chinnadhurai Sankar, Zhou Yu and Alborz Geramifard .....	21
<i>Individual Interaction Styles: Evidence from a Spoken Chat Corpus</i> Nigel Ward .....	27
<i>Evaluation of In-Person Counseling Strategies To Develop Physical Activity Chatbot for Women</i> Kai-Hui Liang, Patrick Lange, Yoo Jung Oh, Jingwen Zhang, Yoshimi Fukuoka and Zhou Yu ..	32
<i>Improving Named Entity Recognition in Spoken Dialog Systems by Context and Speech Pattern Modeling</i> Minh Nguyen and Zhou Yu .....	45
<i>SoDA: On-device Conversational Slot Extraction</i> Sujith Ravi and Zornitsa Kozareva .....	56
<i>Getting to Production with Few-shot Natural Language Generation Models</i> Peyman Heidari, Arash Einolghozati, Shashank Jain, Soumya Batra, Lee Callender, Ankit Arun, Shawn Mei, Sonal Gupta, Pinar Donmez, Vikas Bhardwaj, Anuj Kumar and Michael White .....	66
<i>ARTA: Collection and Classification of Ambiguous Requests and Thoughtful Actions</i> Shohei Tanaka, Koichiro Yoshino, Katsuhito Sudoh and Satoshi Nakamura .....	77
<i>Integrated taxonomy of errors in chat-oriented dialogue systems</i> Ryuichiro Higashinaka, Masahiro Araki, Hiroshi Tsukahara and Masahiro Mizukami .....	89
<i>Effective Social Chatbot Strategies for Increasing User Initiative</i> Amelia Hardy, Ashwin Paranjape and Christopher Manning .....	99
<i>Generative Conversational Networks</i> Alexandros Papangelis, Karthik Gopalakrishnan, Aishwarya Padmakumar, Seokhwan Kim, Gokhan Tur and Dilek Hakkani-Tur .....	111
<i>Commonsense-Focused Dialogues for Response Generation: An Empirical Study</i> Pei Zhou, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu and Dilek Hakkani-Tur .....	121
<i>Velocidapter: Task-oriented Dialogue Comprehension Modeling Pairing Synthetic Text Generation with Domain Adaptation</i> Ibrahim Taha Aksu, Zhengyuan Liu, Min-Yen Kan and Nancy Chen .....	133
<i>An Analysis of State-of-the-Art Models for Situated Interactive MultiModal Conversations (SIMMC)</i> Satwik Kottur, Paul Crook, Seungwhan Moon, Ahmad Beirami, Eunjoon Cho, Rajen Subba and Alborz Geramifard .....	144
<i>A Simple yet Effective Method for Sentence Ordering</i> Aili Shen and Timothy Baldwin .....	154

<i>Topic Shift Detection for Mixed Initiative Response</i>	
Rachna Konigari, Saurabh Ramola, Vijay Vardhan Alluri and Manish Shrivastava . . . . .	161
<i>Improving Unsupervised Dialogue Topic Segmentation with Utterance-Pair Coherence Scoring</i>	
Linzi Xing and Giuseppe Carenini . . . . .	167
<i>Fundamental Exploration of Evaluation Metrics for Persona Characteristics of Text Utterances</i>	
Chiaki Miyazaki, Saya Kanno, Makoto Yoda, Junya Ono and Hiromi Wakaki . . . . .	178
<i>Multi-Referenced Training for Dialogue Response Generation</i>	
Tianyu Zhao and Tatsuya Kawahara . . . . .	190
<i>Contrastive Response Pairs for Automatic Evaluation of Non-task-oriented Neural Conversational Models</i>	
Koshiro Okano, Yu Suzuki, Masaya Kawamura, Tsuneo Kato, Akihiro Tamura and Jianming Wu	202
<i>How does BERT process disfluency?</i>	
Ye Tian, Tim Nieradzick, Sepehr Jalali and Da-shan Shiu . . . . .	208
<i>Hi-DST: A Hierarchical Approach for Scalable and Extensible Dialogue State Tracking</i>	
Suvodip Dey and Maunendra Sankar Desarkar . . . . .	218
<i>Dialogue State Tracking with Multi-Level Fusion of Predicted Dialogue States and Conversations</i>	
Jingyao Zhou, Haipang Wu, Zehao Lin, Guodun Li and Yin Zhang . . . . .	228
<i>Recent Neural Methods on Dialogue State Tracking for Task-Oriented Dialogue Systems: A Survey</i>	
Vevake Balaraman, Seyedmostafa Sheikhalishahi and Bernardo Magnini . . . . .	239
<i>Scikit-talk: A toolkit for processing real-world conversational speech data</i>	
Andreas Liesenfeld, Gabor Parti and Chu-Ren Huang . . . . .	252
<i>ERICA: An Empathetic Android Companion for Covid-19 Quarantine</i>	
Etsuko Ishii, Genta Indra Winata, Samuel Cahyawijaya, Divesh Lala, Tatsuya Kawahara and Pascale Fung . . . . .	257
<i>A multi-party attentive listening robot which stimulates involvement from side participants</i>	
Koji Inoue, Hiromi Sakamoto, Kenta Yamamoto, Divesh Lala and Tatsuya Kawahara . . . . .	261
<i>A Cloud-based User-Centered Time-Offset Interaction Application</i>	
Alberto Chierici, Tyece Kiana Fredorcia Hensley, Wahib Kamran, Kertu Koss, Armaan Agrawal, Erin Meekhof, Goffredo Puccetti and Nizar Habash . . . . .	265
<i>Telling Stories through Multi-User Dialogue by Modeling Character Relations</i>	
Wai Man Si, Prithviraj Ammanabrolu and Mark Riedl . . . . .	269
<i>Summarizing Behavioral Change Goals from SMS Exchanges to Support Health Coaches</i>	
Itika Gupta, Barbara Di Eugenio, Brian D. Ziebart, Bing Liu, Ben S. Gerber and Lisa K. Sharp	276
<i>Rare-Class Dialogue Act Tagging for Alzheimer’s Disease Diagnosis</i>	
Shamila Nasreen, Julian Hough and Matthew Purver . . . . .	290
<i>CIDER: Commonsense Inference for Dialogue Explanation and Reasoning</i>	
Deepanway Ghosal, Pengfei Hong, Siqi Shen, Navonil Majumder, Rada Mihalcea and Soujanya Poria . . . . .	301

<i>Where Are We in Discourse Relation Recognition?</i>	
Katherine Atwell, Junyi Jessy Li and Malihe Alikhani . . . . .	314
<i>Annotation Inconsistency and Entity Bias in MultiWOZ</i>	
Kun Qian, Ahmad Beirami, Zhouhan Lin, Ankita De, Alborz Geramifard, Zhou Yu and Chinnadhurai Sankar . . . . .	326
<i>On the Need for Thoughtful Data Collection for Multi-Party Dialogue: A Survey of Available Corpora and Collection Methods</i>	
Khyati Mahajan and Samira Shaikh . . . . .	338
<i>How Should Agents Ask Questions For Situated Learning? An Annotated Dialogue Corpus</i>	
Felix Gervits, Antonio Roque, Gordon Briggs, Matthias Scheutz and Matthew Marge . . . . .	353
<i>How Will I Argue? A Dataset for Evaluating Recommender Systems for Argumentations</i>	
Markus Brenneis, Maike Behrendt and Stefan Harmeling . . . . .	360
<i>From Argument Search to Argumentative Dialogue: A Topic-independent Approach to Argument Acquisition for Dialogue Systems</i>	
Niklas Rach, Carolin Schindler, Isabel Feustel, Johannes Daxenberger, Wolfgang Minker and Stefan Ultes . . . . .	368
<i>What to Fact-Check: Guiding Check-Worthy Information Detection in News Articles through Argumentative Discourse Structure</i>	
Tariq Alhindi, Brennan McManus and Smaranda Muresan . . . . .	380
<i>How "open" are the conversations with open-domain chatbots? A proposal for Speech Event based evaluation</i>	
A. Seza Dođruöz and Gabriel Skantze . . . . .	392
<i>Blending Task Success and User Satisfaction: Analysis of Learned Dialogue Behaviour with Multiple Rewards</i>	
Stefan Ultes and Wolfgang Maier . . . . .	403
<i>Diversity as a By-Product: Goal-oriented Language Generation Leads to Linguistic Variation</i>	
Simeon Schüz, Ting Han and Sina Zarriß . . . . .	411
<i>DTAFA: Decoupled Training Architecture for Efficient FAQ Retrieval</i>	
Haytham Assem, Sourav Dutta and Edward Burgin . . . . .	423
<i>Projection of Turn Completion in Incremental Spoken Dialogue Systems</i>	
Erik Ekstedt and Gabriel Skantze . . . . .	431
<i>A Task-Oriented Dialogue Architecture via Transformer Neural Language Models and Symbolic Injection</i>	
Oscar J Romero, Antian Wang, John Zimmerman, Aaron Steinfeld and Anthony Tomasic . . . . .	438
<i>Domain-independent User Simulation with Transformers for Task-oriented Dialogue Systems</i>	
Hsien-chin Lin, Nurul Lubis, Songbo Hu, Carel van Niekerk, Christian Geishausser, Michael Heck, Shutong Feng and Milica Gasic . . . . .	445
<i>A Practical 2-step Approach to Assist Enterprise Question-Answering Live Chat</i>	
Ling-Yen Liao and Tarec Fares . . . . .	457

<i>A Brief Study on the Effects of Training Generative Dialogue Models with a Semantic loss</i> Prasanna Parthasarathi, Mohamed Abdelsalam, Sarath Chandar and Joelle Pineau .....	469
<i>Do Encoder Representations of Generative Dialogue Models have sufficient summary of the Information about the task ?</i> Prasanna Parthasarathi, Joelle Pineau and Sarath Chandar .....	477
<i>GenSF: Simultaneous Adaptation of Generative Pre-trained Models and Slot Filling</i> Shikib Mehri and Maxine Eskenazi .....	489
<i>Schema-Guided Paradigm for Zero-Shot Dialog</i> Shikib Mehri and Maxine Eskenazi .....	499
<i>Coreference-Aware Dialogue Summarization</i> Zhengyuan Liu, Ke Shi and Nancy Chen .....	509
<i>Weakly Supervised Extractive Summarization with Attention</i> Yingying Zhuang, Yichao Lu and Simi Wang .....	520
<i>Incremental temporal summarization in multi-party meetings</i> Ramesh Manuvinaurike, Saurav Sahay, Wenda Chen and Lama Nachman .....	530
<i>Mitigating Topic Bias when Detecting Decisions in Dialogue</i> Mladen Karan, Prashant Khare, Patrick Healey and Matthew Purver .....	542
<i>Assessing Political Prudence of Open-domain Chatbots</i> Yejin Bang, Nayeon Lee, Etsuko Ishii, Andrea Madotto and Pascale Fung .....	548
<i>Large-Scale Quantitative Evaluation of Dialogue Agents' Response Strategies against Offensive Users</i> Haojun Li, Dilara Soylu and Christopher Manning .....	556



# Conference Program

All times are shown in Singapore local time (GMT+8).

**July 29, 2021**

**20:00–20:30** *BREAKOUT*

**20:30–21:00** *Opening Ceremony*

21:00–22:15 *Keynote 1: Dialog with Robots: Perceptually Grounded Communication with Life-long Learning*  
Raymond J. Mooney

**22:30–03:00** *Special Session: Summarization of Dialogues and Multi-Party Meetings (Summ-Dial)*

**22:00–00:15** *Special Session: Safety for E2E Conversational AI (SafeConvAI)*

**July 30, 2021**

**12:00–13:00** **Paper Session P1**

*Understanding and predicting user dissatisfaction in a neural generative chatbot*  
Abigail See and Christopher Manning

*Towards Continuous Estimation of Dissatisfaction in Spoken Dialog*  
Nigel Ward, Jonathan E. Avila and Aaron M. Alarcon

*DialogStitch: Synthetic Deeper and Multi-Context Task-Oriented Dialogs*  
Satwik Kottur, Chinnadhurai Sankar, Zhou Yu and Alborz Geramifard

*Individual Interaction Styles: Evidence from a Spoken Chat Corpus*  
Nigel Ward

*Evaluation of In-Person Counseling Strategies To Develop Physical Activity Chatbot for Women*

Kai-Hui Liang, Patrick Lange, Yoo Jung Oh, Jingwen Zhang, Yoshimi Fukuoka and Zhou Yu

**July 30, 2021 (continued)**

*Improving Named Entity Recognition in Spoken Dialog Systems by Context and Speech Pattern Modeling*

Minh Nguyen and Zhou Yu

*SoDA: On-device Conversational Slot Extraction*

Sujith Ravi and Zornitsa Kozareva

*Getting to Production with Few-shot Natural Language Generation Models*

Peyman Heidari, Arash Einolghozati, Shashank Jain, Soumya Batra, Lee Callender, Ankit Arun, Shawn Mei, Sonal Gupta, Pinar Donmez, Vikas Bhardwaj, Anuj Kumar and Michael White

**13:00–14:30 Paper Session P2**

*ARTA: Collection and Classification of Ambiguous Requests and Thoughtful Actions*

Shohei Tanaka, Koichiro Yoshino, Katsuhito Sudoh and Satoshi Nakamura

*Integrated taxonomy of errors in chat-oriented dialogue systems*

Ryuichiro Higashinaka, Masahiro Araki, Hiroshi Tsukahara and Masahiro Mizukami

*Effective Social Chatbot Strategies for Increasing User Initiative*

Amelia Hardy, Ashwin Paranjape and Christopher Manning

*Generative Conversational Networks*

Alexandros Papangelis, Karthik Gopalakrishnan, Aishwarya Padmakumar, Seokhwan Kim, Gokhan Tur and Dilek Hakkani-Tur

*Commonsense-Focused Dialogues for Response Generation: An Empirical Study*

Pei Zhou, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu and Dilek Hakkani-Tur

*Velocidapter: Task-oriented Dialogue Comprehension Modeling Pairing Synthetic Text Generation with Domain Adaptation*

Ibrahim Taha Aksu, Zhengyuan Liu, Min-Yen Kan and Nancy Chen

*An Analysis of State-of-the-Art Models for Situated Interactive MultiModal Conversations (SIMMC)*

Satwik Kottur, Paul Crook, Seungwhan Moon, Ahmad Beirami, Eunjoon Cho, Rajen Subba and Alborz Geramifard

*A Simple yet Effective Method for Sentence Ordering*

Aili Shen and Timothy Baldwin

**July 30, 2021 (continued)**

*Topic Shift Detection for Mixed Initiative Response*

Rachna Konigari, Saurabh Ramola, Vijay Vardhan Alluri and Manish Shrivastava

*Improving Unsupervised Dialogue Topic Segmentation with Utterance-Pair Coherence Scoring*

Linzi Xing and Giuseppe Carenini

**14:30–15:00 BREAKOUT**

**15:00–16:00 Paper Session P3**

*Fundamental Exploration of Evaluation Metrics for Persona Characteristics of Text Utterances*

Chiaki Miyazaki, Saya Kanno, Makoto Yoda, Junya Ono and Hiromi Wakaki

*Multi-Referenced Training for Dialogue Response Generation*

Tianyu Zhao and Tatsuya Kawahara

*Contrastive Response Pairs for Automatic Evaluation of Non-task-oriented Neural Conversational Models*

Koshiro Okano, Yu Suzuki, Masaya Kawamura, Tsuneo Kato, Akihiro Tamura and Jianming Wu

*How does BERT process disfluency?*

Ye Tian, Tim Nieradzic, Sepehr Jalali and Da-shan Shiu

*Hi-DST: A Hierarchical Approach for Scalable and Extensible Dialogue State Tracking*

Suvodip Dey and Maunendra Sankar Desarkar

*Dialogue State Tracking with Multi-Level Fusion of Predicted Dialogue States and Conversations*

Jingyao Zhou, Haipang Wu, Zehao Lin, Guodun Li and Yin Zhang

*Recent Neural Methods on Dialogue State Tracking for Task-Oriented Dialogue Systems: A Survey*

Vevake Balaraman, Seyedmostafa Sheikhalishahi and Bernardo Magnini

**July 30, 2021 (continued)**

**16:00–17:00 Demo Session**

*Scikit-talk: A toolkit for processing real-world conversational speech data*

Andreas Liesenfeld, Gabor Parti and Chu-Ren Huang

*ERICA: An Empathetic Android Companion for Covid-19 Quarantine*

Etsuko Ishii, Genta Indra Winata, Samuel Cahyawijaya, Divesh Lala, Tatsuya Kawahara and Pascale Fung

*A multi-party attentive listening robot which stimulates involvement from side participants*

Koji Inoue, Hiromi Sakamoto, Kenta Yamamoto, Divesh Lala and Tatsuya Kawahara

*A Cloud-based User-Centered Time-Offset Interaction Application*

Alberto Chierici, Tyece Kiana Fredorcia Hensley, Wahib Kamran, Kertu Koss, Armaan Agrawal, Erin Meekhof, Goffredo Puccetti and Nizar Habash

**17:30–19:00 Panel**

**19:00–20:00 Virtual Tour (ALL) and Dinner (Physical only)**

**20:00–21:00 Sponsor Session SPSI**

21:00–22:15 *Keynote 2: A journey from ML and NNs to NLP and Beyond: Just more of the same isn't enough?*

Jason Weston

July 30, 2021 (continued)

23:00–00:00 Paper Session P4

*Telling Stories through Multi-User Dialogue by Modeling Character Relations*

Wai Man Si, Prithviraj Ammanabrolu and Mark Riedl

*Summarizing Behavioral Change Goals from SMS Exchanges to Support Health Coaches*

Itika Gupta, Barbara Di Eugenio, Brian D. Ziebart, Bing Liu, Ben S. Gerber and Lisa K. Sharp

*Rare-Class Dialogue Act Tagging for Alzheimer’s Disease Diagnosis*

Shamila Nasreen, Julian Hough and Matthew Purver

*CIDER: Commonsense Inference for Dialogue Explanation and Reasoning*

Deepanway Ghosal, Pengfei Hong, Siqi Shen, Navonil Majumder, Rada Mihalcea and Soujanya Poria

*Where Are We in Discourse Relation Recognition?*

Katherine Atwell, Junyi Jessy Li and Malihe Alikhani

*Annotation Inconsistency and Entity Bias in MultiWOZ*

Kun Qian, Ahmad Beirami, Zhouhan Lin, Ankita De, Alborz Geramifard, Zhou Yu and Chinnadhurai Sankar

*On the Need for Thoughtful Data Collection for Multi-Party Dialogue: A Survey of Available Corpora and Collection Methods*

Khyati Mahajan and Samira Shaikh

*How Should Agents Ask Questions For Situated Learning? An Annotated Dialogue Corpus*

Felix Gervits, Antonio Roque, Gordon Briggs, Matthias Scheutz and Matthew Marge

**July 31, 2021**

**19:00–20:00** *Sponsor Session SPS2*

**20:00–21:00** **Paper Session P5**

*How Will I Argue? A Dataset for Evaluating Recommender Systems for Argumentations*

Markus Brenneis, Maike Behrendt and Stefan Harmeling

*From Argument Search to Argumentative Dialogue: A Topic-independent Approach to Argument Acquisition for Dialogue Systems*

Niklas Rach, Carolin Schindler, Isabel Feustel, Johannes Daxenberger, Wolfgang Minker and Stefan Ultes

*What to Fact-Check: Guiding Check-Worthy Information Detection in News Articles through Argumentative Discourse Structure*

Tariq Alhindi, Brennan McManus and Smaranda Muresan

*How "open" are the conversations with open-domain chatbots? A proposal for Speech Event based evaluation*

A. Seza Dođruöz and Gabriel Skantze

*Blending Task Success and User Satisfaction: Analysis of Learned Dialogue Behaviour with Multiple Rewards*

Stefan Ultes and Wolfgang Maier

*Diversity as a By-Product: Goal-oriented Language Generation Leads to Linguistic Variation*

Simeon Schüz, Ting Han and Sina Zarriß

*DTAFA: Decoupled Training Architecture for Efficient FAQ Retrieval*

Haytham Assem, Sourav Dutta and Edward Burgin

*Projection of Turn Completion in Incremental Spoken Dialogue Systems*

Erik Ekstedt and Gabriel Skantze

**July 31, 2021 (continued)**

**21:00–22:00 Paper Session P6**

*A Task-Oriented Dialogue Architecture via Transformer Neural Language Models and Symbolic Injection*

Oscar J Romero, Antian Wang, John Zimmerman, Aaron Steinfeld and Anthony Tomasic

*Domain-independent User Simulation with Transformers for Task-oriented Dialogue Systems*

Hsien-chin Lin, Nurul Lubis, Songbo Hu, Carel van Niekerk, Christian Geishauer, Michael Heck, Shutong Feng and Milica Gasic

*A Practical 2-step Approach to Assist Enterprise Question-Answering Live Chat*

Ling-Yen Liao and Tarec Fares

*A Brief Study on the Effects of Training Generative Dialogue Models with a Semantic loss*

Prasanna Parthasarathi, Mohamed Abdelsalam, Sarath Chandar and Joelle Pineau

*Do Encoder Representations of Generative Dialogue Models have sufficient summary of the Information about the task ?*

Prasanna Parthasarathi, Joelle Pineau and Sarath Chandar

*GenSF: Simultaneous Adaptation of Generative Pre-trained Models and Slot Filling*

Shikib Mehri and Maxine Eskenazi

*Schema-Guided Paradigm for Zero-Shot Dialog*

Shikib Mehri and Maxine Eskenazi

**22:00–22:30 BREAKOUT**

22:30–23:45 *Keynote 3: Whom Do We Trust in Dialogue Systems?*

Julia Hirschberg

**00:00–01:00 Business Meeting and Closing Ceremony**

## Special Session: Summarization of Dialogues and Multi-Party Meetings (SummDial)

### 22:30–22:45 *Opening*

22:45–23:30 *Keynote 1: Who discussed what with whom: is meeting summarization a solved problem?*  
Klaus Zechner

### 23:30–23:35 *Break*

23:35–23:55 *Coreference-Aware Dialogue Summarization*  
Zhengyuan Liu, Ke Shi and Nancy Chen

23:55–00:15 *Weakly Supervised Extractive Summarization with Attention*  
Yingying Zhuang, Yichao Lu and Simi Wang

00:15–00:35 *Incremental temporal summarization in multi-party meetings*  
Ramesh Manuvinakurike, Saurav Sahay, Wenda Chen and Lama Nachman

### 00:35–00:45 *Break*

00:45–01:45 *Panel Discussion: Dialogue and Meeting Summarization: Taking Stock and Looking Ahead*  
Ani Nenkova, Klaus Zechner, Diyi Yang, and Chenguang Zhu

### 01:45–01:50 *Break*

01:50–02:10 *Mitigating Topic Bias when Detecting Decisions in Dialogue*  
Mladen Karan, Prashant Khare, Patrick Healey and Matthew Purver

02:10–02:30 *Creating a data set of abstractive summaries of turn-labeled spoken human-computer conversations*  
Iris Hendrickx and Virginia Meijer

02:30–02:50 *Dynamic Sliding Window for Meeting Summarization*  
Zhengyuan Liu and Nancy Chen



**Special Session: Summarization of Dialogues and Multi-Party Meetings (SummDial) (continued)**

**02:50–03:00** *Closing*

**Special Session: Safety for E2E Conversational AI (SafeConvAI)**

22:00–22:10 *Welcome by the organisers*  
Verena Rieser

22:10–22:50 *Keynote*  
Laurence Devillers

**22:50–23:00** *Coffee break*

**23:00–23:30** **Paper presentations**

*Assessing Political Prudence of Open-domain Chatbots*

Yejin Bang, Nayeon Lee, Etsuko Ishii, Andrea Madotto and Pascale Fung

*Large-Scale Quantitative Evaluation of Dialogue Agents' Response Strategies against Offensive Users*

Haojun Li, Dilara Soylu and Christopher Manning

*Panel discussion*

Pascale Fung, Pilar Manchon, Ehud Reiter, Michelle Zhou, Emily Dinan (session chair)



## Keynote Abstracts

### **Keynote 1 - Dialog with Robots: Perceptually Grounded Communication with Lifelong Learning**

Raymond J. Mooney

*The University of Texas at Austin*

#### **Abstract**

Developing robots that can accept instructions from and collaborate with human users is greatly enhanced by an ability to engage in natural language dialog. Unlike most other dialog scenarios, this requires grounding the semantic analysis of language in perception and action in the world. Although deep-learning has greatly enhanced methods for such grounded language understanding, it is difficult to ensure that the data used to train such models covers all of the concepts that a robot might encounter in practice. Therefore, we have developed methods that can continue to learn from dialog with users during ordinary use by acquiring additional targeted training data from the responses to intentionally designed clarification and active learning queries. These methods use reinforcement learning to automatically acquire dialog strategies that support both effective immediate task completion as well as learning that improves future performance. Using both experiments in simulation and with real robots, we have demonstrated that these methods exhibit life-long learning that improves long-term performance.

#### **Biography**

Raymond J. Mooney is a Professor in the Department of Computer Science at the University of Texas at Austin. He received his Ph.D. in 1988 from the University of Illinois at Urbana/Champaign. He is an author of over 180 published research papers, primarily in the areas of machine learning and natural language processing. He was the President of the International Machine Learning Society from 2008-2011, program co-chair for AAAI 2006, general chair for HLT-EMNLP 2005, and co-chair for ICML 1990. He is a Fellow of AAAI, ACM, and ACL and the recipient of the Classic Paper award from AAAI-19 and best paper awards from AAAI-96, KDD-04, ICML-05 and ACL-07.

## **Keynote 2 - A journey from ML & NNs to NLP and Beyond: Just more of the same isn't enough?**

Jason Weston

*Facebook AI & NYU*

### **Abstract**

The first half of the talk will look back on the last two decades of machine learning, neural network and natural language processing research for dialogue, through my personal lens, to discuss the advances that have been made and the circumstances in which they happened — to try to give clues of what we should be working on for the future. The second half will dive deeper into some current first steps in those future directions, in particular trying to fix the problems of neural generative models to enable deeper reasoning with short and long-term coherence, and to ground such dialogue agents to an environment where they can act and learn. We will argue that just scaling up current techniques, while a worthy investigation, will not be enough to solve these problems.

### **Biography**

Jason Weston is a research scientist at Facebook, NY and a Visiting Research Professor at NYU. He earned his PhD in machine learning at Royal Holloway, University of London and at AT&T Research in Red Bank, NJ (advisors: Alex Gammerman, Volodya Vovk and Vladimir Vapnik) in 2000. From 2000 to 2001, he was a researcher at Biowulf technologies. From 2002 to 2003 he was a research scientist at the Max Planck Institute for Biological Cybernetics, Tuebingen, Germany. From 2003 to 2009 he was a research staff member at NEC Labs America, Princeton. From 2009 to 2014 he was a research scientist at Google, NY. His interests lie in statistical machine learning, with a focus on reasoning, memory, perception, interaction and communication. Jason has published over 100 papers, including best paper awards at ICML and ECML, and a Test of Time Award for his work “A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning”, ICML 2008 (with Ronan Collobert). He was part of the YouTube team that won a National Academy of Television Arts & Sciences Emmy Award for Technology and Engineering for Personalized Recommendation Engines for Video Discovery. He was listed as the 16th most influential machine learning scholar at AMiner and one of the top 50 authors in Computer Science in Science.

## **Keynote 3 - Whom Do We Trust in Dialogue Systems?**

Julia Hirschberg

*Columbia University*

### **Abstract**

It is important for computer systems today to encourage user trust: for recommender systems, knowledge-delivery systems, and dialogue systems in general. What aspects of text or speech production do humans tend to trust? It is also important for these systems to be able to identify whether in fact a user does trust them. But producing trusted speech and recognizing user trust are still challenging questions. Our work on trusted and mistrusted speech has produced some useful information about the first issue, exploring the types of lexical and acoustic-prosodic features in human speech that listeners tend to trust or to mistrust. Using the very large Columbia Cross-cultural Deception Corpus we created to detect truth vs. lie, we created a LieCatcher game to crowd-source a project on trusted vs. mistrusted speech from multiple raters listening to question responses and rating them as true or false. We present results on the types of speech raters trusted or did not trust and their reasoning behind their answers. We then describe ongoing research on the second issue: How do we determine whether a user trusts the system and do aspects of their speech reveal useful information?

### **Biography**

Julia Hirschberg is Percy K. and Vida L. W. Hudson Professor of Computer Science at Columbia University. She previously worked at Bell Laboratories and AT&T Labs on text-to-speech synthesis (TTS) and created their first HCI Research Department. She is a fellow of AAAI, ISCA, ACL, ACM, and IEEE, and a member of the NAE, the American Academy of Arts and Sciences, and the American Philosophical Society, and has received the IEEE James L. Flanagan Speech and Audio Processing Award, the ISCA Medal for Scientific Achievement and the ISCA Special Service Medal. She studies speech and NLP, currently TTS; deceptive, trusted, emotional, and charismatic speech; false information and intent on social media; multimodal humor; and radicalization. She has worked for diversity for many years at AT&T and Columbia.

## **SummDial Keynote - Who discussed what with whom: is meeting summarization a solved problem?**

Klaus Zechner

*Educational Testing Service, United States*

### **Abstract**

While creating audio and video records of multi-party meetings has become easier than ever in recent years, obtaining access to the key contents or a summary of a meeting is non-trivial. In this talk, I will first provide an overview of the main differences between multi-party meetings and news articles – the prototypical domain for most research on summarization so far. In the second part of the talk, a few example approaches to meeting summarization will be presented and discussed, spanning from early research to late-breaking system papers. Finally, I will conclude with thoughts about the current state-of-the-art of the field of meeting summarization and open issues that still need to be addressed by the research community.

### **Biography**

Klaus Zechner received his Ph.D. from Carnegie Mellon University in 2001 for research on automated speech summarization. This work was published at SIGIR-2001 and in *Computational Linguistics* (2002). Klaus Zechner is now a Senior Research Scientist in the Natural Language Processing Lab in the Research and Development Division of Educational Testing Service (ETS) in Princeton, New Jersey, USA. Since joining ETS in 2002, he has been pioneering research and development of technologies for automated scoring of non-native speech, leading large R&D projects dedicated to the continuous improvement of automated speech scoring technology. He holds more than 20 patents on technology related to SpeechRater®, an automated speech scoring system he and his team have been developing at ETS. SpeechRater is currently used operationally as sole score for the TOEFL®Practice Online (TPO) Speaking assessment and, in a hybrid scoring approach, also for TOEFL iBT Speaking. Klaus Zechner authored more than 80 peer-reviewed publications in journals, book chapters, conference and workshop proceedings, and research reports. He also edited a book on automated speaking assessment that was published by Routledge in 2019; it provides an overview of the current state-of-the-art in automated speech scoring of spontaneous non-native speech.

## **SafeConvAI Keynote - Emotional manipulation of chatbots: the nudge**

Laurence Devillers

*Sorbonne University - CNRS-LISN (Saclay)*

### **Abstract**

While creating audio and video records of multi-party meetings has become easier than ever in recent years, obtaining access to the key contents or a summary of a meeting is non-trivial. In this talk, I will first provide an overview of the main differences between multi-party meetings and news articles – the prototypical domain for most research on summarization so far. In the second part of the talk, a few example approaches to meeting summarization will be presented and discussed, spanning from early research to late-breaking system papers. Finally, I will conclude with thoughts about the current state-of-the-art of the field of meeting summarization and open issues that still need to be addressed by the research community.

### **Biography**

Laurence Devillers is a full Professor of Artificial Intelligence at Sorbonne University and heads the team of research “Affective and social dimensions in Spoken interaction with (ro)bots: ethical issues” at CNRS-LISN (Saclay). Since 2020, she heads the interdisciplinary Chair on Artificial Intelligence HUMANMAINE: HUMAN-MACHINE Affective INTERACTION & ETHICS (2020-24) at CNRS. Her topics of research are Human-Machine Co-adaptation: from the modeling of emotions and human-robot dialogue to the ethical impacts for society and the risks and benefits of AI notably for vulnerable people. She is a member of National Comity Pilot on Ethics of Numeric (CNPEN) working on conversational Agents, social robots, AI and Ethics. She is now an expert member of the GPAI on “the future of work” since June 2020 (international group). In March 2020, she wrote the book “Les robots émotionnels” (Ed. L’Observatoire) and in March 2017 “Des Robots et des Hommes: mythes, fantasmes et réalité” (Ed. Plon) for explaining the urgency of building Social and Affective Robotic/AI Systems with Ethics by design.

