

DialDoc 2023

**Proceedings of the Third DialDoc Workshop on
Document-grounded Dialogue and Conversational Question
Answering**

Proceedings of the Workshop

July 13, 2023

The DialDoc organizers gratefully acknowledge the support from the following sponsors.

Gold



©2023 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-959429-98-2

Introduction

Welcome to the Third Workshop on Document-grounded Dialogue and Conversational Question Answering (DialDoc), co-located with ACL 2023.

The DialDoc workshop aims to address the challenges of Document-Grounded Dialogue and Conversational Question Answering. In today's world, where a vast amount of content is generated across various mediums, it becomes crucial to not only make this content accessible to users through conversational interfaces but also ensure that the responses provided by models are accurate and grounded in reliable knowledge sources.

In our third workshop, we are particularly interested in exploring the theme of Factual Consistency. With the recent advancements in large language models, a significant concern arises when these systems generate responses that contain factual inconsistencies compared to external sources. This issue has implications on user trust and safety. We aim to highlight important community efforts that address the challenges associated with factual consistency, including but not limited to automatic evaluation methods, human evaluation, modeling techniques, and datasets.

The Shared Task competition primarily focuses on developing goal-oriented document-grounded dialogue systems in a multilingual setting. These systems allow users to interactively query domain-specific information based on provided documents. The task of querying document knowledge through conversational systems has gained considerable attention from both research and industrial communities due to its various applications. While previous Shared Tasks organized by the First and Second DialDoc Workshops focused on English document-grounded dialogue systems, other languages have been less explored. As a result, large communities of users are unable to access automated services and information. In order to bridge this gap, the Third DialDoc Workshop introduces a shared task that involves documents and dialogues in diverse languages. The aim is to encourage researchers to explore effective solutions for two key challenges: (1) transferring a DGD model from a high-resource language to a low-resource language, and (2) developing a DGD model capable of providing multilingual responses given multilingual documents. To evaluate the performance of response generation, the workshop adopts token-level F1, SacreBleu, and Rouge-L metrics. The score is calculated based on the sum of these metrics. A total of 71 teams participated in the Dev Phase, and for the final Test Phase, 29 teams submitted their models to the leaderboards. Many submissions have significantly outperformed the baseline, with the best-performing system achieving a score of 215.4 compared to the baseline's score of 156.0.

The workshop received a total of 19 submissions, featuring 18 paper presentations in either poster or oral format. Additionally, we are privileged to have invited talks from Greg Durrett, Hannaneh Hajishirz, Xiang Ren, and Rui Yan.

We would like to express our gratitude to all those who have contributed to the success of this workshop. Our thanks go to the authors for their valuable paper submissions, the teams for their participation in the Shared Task competition, the program committee members for their significant contributions, and the ACL workshop co-chairs for their guidance. We are also grateful to our esteemed invited speakers. Special appreciation is extended to Alibaba DAMO Academy for their sponsorship of the rewards for the Shared Task competition.

Organizing Committee

Program Chairs

Roe Aharoni, Google Research
Nouha Dziri, Allen Institute for AI
Song Feng, AWS AI Labs
Yongbin Li, DAMO Academy, Alibaba Group
Yu Li, Columbia University
Hui Wan, IBM Research AI

Shared Task Organizers

Haiyang Yu, DAMO Academy, Alibaba Group
Cam-Tu Nguyen, Nanjing University
Bowen Yu, DAMO Academy, Alibaba Group
Haomin Fu, Nanjing University
Yeqin Zhang, Nanjing University

Program Committee

Reviewer

Ankita Gupta
Chang Gao
Chulaka Gunasekara
Nico Daheim
Daniele Bonadiman
Etsuko Ishii
Hao Fang
Han He
Haomin Fu
Ehsan Kamaloo
Kun Qian
Anh-Cuong Le
Hoang-Quynh Le
Min Yang
Nguyen Cam-Tu
Pengshan Cai
Srijan Bansal
Sam Davidson
Wanyu Du
Xiangkun Hu
Xing Han Lu
Bowen Yu
Yeqin Zhang
Zichao Li
Ankita Gupta
Chulaka Gunasekara
Ehsan Kamaloo
Etsuko Ishii
Hao Fang

Keynote Talk: Assessing LLM Faithfulness: Lessons from Political Fact-checking

Greg Durrett

The University of Texas at Austin

Bio: Greg Durrett is an assistant professor of Computer Science at UT Austin. His research focuses on techniques for accessing and reasoning about knowledge in text. Large language models (LLMs) like ChatGPT and GPT-4 have dramatically advanced the frontiers in this area; currently his team is looking at where these systems succeed and fail and how to enhance their capabilities, particularly via systems that use LLMs as primitives. He is a 2023 Sloan Research Fellow and a recipient of a 2022 NSF CAREER award, among other grants from agencies including the NSF, Open Philanthropy, DARPA, Salesforce, and Amazon. He completed his Ph.D. at UC Berkeley where he was advised by Dan Klein, and he was previously a research scientist at Semantic Machines.

Keynote Talk: Reflex or Reflect: When Do Language Tasks Need Slow Reasoning?

Xiang Ren

University of Southern California

Bio: Xiang Ren is an assistant professor at USC Computer Science Department and a Research Team Lead at USC ISI. He is the director of Intelligence and Knowledge Discovery (INK) Research Lab, the Information Director of ACM SIGKDD, and member of USC Machine Learning Center. Priorly, he was a research scholar at Stanford University, and received his Ph.D. in Computer Science from University of Illinois Urbana-Champaign. Dr. Ren's research focuses on developing label-efficient, prior-informed computational methods that extract machine-actionable knowledge from natural-language data, as well as performing neural-symbolic reasoning over heterogeneous data. His research leads to a book and over 50 publications, was covered in over 10 conference tutorials, and received awards including faculty research awards from Google, Amazon, JP Morgan, Sony and Snapchat, ACM SIGKDD Dissertation Award, The Web Conference Best Paper award honorable mention, and David J. Kuck Outstanding Thesis Award. He was named Forbes' Asia 30 Under 30 in 2019.

Keynote Talk: Improved Factual Precision in Long-form Text Generation with Fine-grained Evaluation and Feedback

Hannaneh Hajishirz
University of Washington

Bio: Hanna Hajishirzi is a Torode Family Associate Professor in the Paul G. Allen School of Computer Science Engineering at the University of Washington and a Senior Research Manager at the Allen Institute for AI. Her research spans different areas in NLP and AI, focusing on developing general-purpose machine learning algorithms that can solve diverse NLP tasks. Applications for these algorithms include question answering, representation learning, green AI, knowledge extraction, and conversational dialogue. Honors include the NSF CAREER Award, Sloan Fellowship, Allen Distinguished Investigator Award, Intel rising star award, best paper and honorable mention awards, and several industry research faculty awards. Hanna received her PhD from University of Illinois and spent a year as a postdoc at Disney Research and CMU.

Keynote Talk: Recent Progress of Conversational AI in the Open Domain

Rui Yan

Renmin University of China

Bio: Rui Yan is an associate professor with tenure at Gaoling School of Artificial Intelligence, Renmin University of China. He was selected as a young scientist at Beijing Academy of Artificial Intelligence (BAAI) and a startrack young fellow of Microsoft Research Asia (MSRA). Till now he has published more than 100 highly peer-reviewed publications with more than 10,000 citations. He regularly served as an area chair/senior PC member for top-tier international conferences. He has been invited to give tutorial talks for these conferences as well.

Table of Contents

<i>Cross-lingual Data Augmentation for Document-grounded Dialog Systems in Low Resource Languages</i> Qi Gou, Zehua Xia and Wenzhe Du	1
<i>MoQA: Benchmarking Multi-Type Open-Domain Question Answering</i> Howard Yen, Tianyu Gao, Jinhyuk Lee and Danqi Chen	8
<i>Exploration of multilingual prompts in document-grounded dialogue</i> Xiaocheng Zhang, Huang Yong Qing and Fu Xue Lin	30
<i>Position Matters! Empirical Study of Order Effect in Knowledge-grounded Dialogue</i> Hsuan Su, Shachi H. Kumar, Sahisnu Mazumder, Wenda Chen, Ramesh Manuvinakurike, Eda Okur, Saurav Sahay, Lama Nachman, Shang-Tse Chen and Hung-yi Lee	36
<i>Enhancing Multilingual Document-Grounded Dialogue Using Cascaded Prompt-Based Post-Training Models</i> Jun Liu, Shuang Cheng, Zineng Zhou, Yang Gu, Jian Ye and Haiyong Luo	44
<i>Enhanced Training Methods for Multiple Languages</i> Hai Li and Yang Li	52
<i>SLDT: Sequential Latent Document Transformer for Multilingual Document-based Dialogue</i> Zhanyu Ma, Zeming Liu and Jian Ye	57
<i>A Dialogue System for Assessing Activities of Daily Living: Improving Consistency with Grounded Knowledge</i> Zhecheng Sheng, Raymond Lewis Finzel, lucke096@umn.edu lucke096@umn.edu, gahmx008@umn.edu gahmx008@umn.edu, Maria Gini and Serguei V. S. Pakhomov	68
<i>C-PMI: Conditional Pointwise Mutual Information for Turn-level Dialogue Evaluation</i> Liliang Ren, Mankeerat Sidhu, Qi Zeng, Revanth Gangi Reddy, Heng Ji and ChengXiang Zhai	80
<i>ConvRGX: Recognition, Generation, and Extraction for Self-trained Conversational Question Answering</i> Tianhua Zhang, Liping Tang, Wei Fang, Hongyin Luo, Xixin Wu, Helen M. Meng and James R. Glass	86
<i>Language-Agnostic Transformers and Assessing ChatGPT-Based Query Rewriting for Multilingual Document-Grounded QA</i> Srinivas Gowriraj, Soham Dinesh Tiwari, Mitali Potnis, Srijan Bansal, Teruko Mitamura and Eric Nyberg	101
<i>Follow the Knowledge: Structural Biases and Artefacts in Knowledge Grounded Dialog Datasets</i> Ehsan Lotfi, Maxime De Bruyn, jeska.buhmann@uantwerpen.be jeska.buhmann@uantwerpen.be and Walter Daelemans	109

Program

Thursday, July 13, 2023

- 08:55 - 09:00 *Opening Remarks*
- 09:00 - 09:35 *Invited talk I by Hannaneh Hajishirz*
- 09:35 - 09:55 *Paper Presentation - Follow the Knowledge: Structural Biases and Artefacts in Knowledge Grounded Dialog Datasets*
- 09:55 - 10:15 *Paper Presentation - Revisiting Sentence Union Generation as a Testbed for Text Consolidation*
- 10:15 - 10:30 *Paper Presentation - Graph-Guided Unsupervised Knowledge Identification for Dialogue Agents*
- 10:30 - 10:50 *Coffee Break*
- 10:50 - 11:25 *Invited talk II by Rui Yan*
- 11:25 - 12:30 *Lightning Talks*
- 12:30 - 13:30 *Lunch Break*
- 13:30 - 14:00 *Shared Task Prizes*
- 14:00 - 14:35 *Invited talk III by Xiang Ren*
- 14:35 - 14:50 *Coffee Break*
- 14:50 - 15:10 *Paper Presentation - AlignScore: Evaluating Factual Consistency with A Unified Alignment Function*
- 15:10 - 15:30 *Paper Presentation - A Dialogue System for Assessing Activities of Daily Living: Improving Consistency with Grounded Knowledge*
- 15:30 - 15:50 *Paper Presentation - Ontologically Faithful Generation of Non-Player Character Dialogues*
- 15:50 - 16:10 *Paper Presentation - MoQA: Benchmarking Multi-Type Open-Domain Question Answering*

Thursday, July 13, 2023 (continued)

16:10 - 16:45 *Invited talk IV by Greg Durrett*

16:45 - 16:50 *Ending Remark*